

## STIC-Biotech/ChemLib

---

**From:** Slobodyansky, Elizabeth  
**Sent:** Thursday, June 27, 2002 11:35 AM  
**T :** STIC-Biotech/ChemLib  
**Subject:** 09/801,852

Please search for case 09/801,852:



SEQ ID NOs: 1 and 3 against commercial ~~and interference~~ databases.

Thank you.

*Elizabeth Slobodyansky, PhD*

Primary Examiner

Art Unit 1652  
CM1 10D11  
703-306-3222

**mail box 10C01**

POINT OF CONTACT:  
PAUL SCHULWITZ  
TECHNICAL INFO. SPECIALIST  
CM1 6B06 TEL. (703) 305-1954

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: 6/28  
Date Completed: 7/1  
Searcher Prep/Review: 10  
Clerical: \_\_\_\_\_  
Online time: 10

TYPE OF SEARCH:  
NA Sequences: 2  
AA Sequences: \_\_\_\_\_  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST(where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 29, 2002, 15:06:19 ; Search time 135.69 Seconds  
(without alignments)  
4018.766 Million cell updates/sec

Title: US-09-801-852A-3  
Perfect score: 2220  
Sequence: 1 ggaaccacccgaactgtc.....agacgttggaagtgcgaaa 2220

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA: \*  
1: /cgn2.6/prodata/2/lna/5A-COMB.seq: \*  
2: /cgn2.6/prodata/2/lna/5B-COMB.seq: \*  
3: /cgn2.6/prodata/2/lna/6A-COMB.seq: \*  
4: /cgn2.6/prodata/2/lna/6B-COMB.seq: \*  
5: /cgn2.6/prodata/2/lna/PCTUS-COMB.seq: \*  
6: /cgn2.6/prodata/2/lna/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2220	100.0	2220	2	US-08-932-376A-3
2	1375	61.9	1738	2	US-08-932-376A-1
3	64.4	2.9	1635	2	US-08-776-210-5
4	64.2	2.9	7218	1	US-08-232-463-14
5	62.8	2.8	1828	3	US-08-362-525-11
6	53.4	2.4	1650	2	US-08-776-210-6
7	53.2	2.4	2539	3	US-09-000-016-3
8	53.2	2.4	2809	3	US-09-000-016-1
9	51	2.3	1505	1	US-07-915-246-1
10	49.2	2.2	20235	3	US-07-642-734C-3
11	49.2	2.2	20235	3	US-08-439-009A-3
12	49	2.2	2064	4	US-08-343-428-1
13	48.4	2.2	6085	4	US-09-029-603-4
14	48.2	2.2	1931	2	US-09-130-114-2
15	47.6	2.1	1965	4	US-09-178-252-26
16	47.4	2.1	3196	4	US-09-704-449-1
17	47	2.1	1485	4	US-09-372-422A-39
18	47	2.1	44377	2	US-08-804-227C-7
19	47	2.1	44377	2	US-08-804-198-1
20	46.4	2.1	320	4	US-09-165-264-7
21	45.8	2.1	1142	3	US-08-765-743-1
22	45	2.0	3300	2	US-08-194-290-6
23	45	2.0	3300	2	US-08-614-377A-6
24	45	2.0	3300	4	US-09-142-648B-6
25	45	2.0	8438	1	US-07-945-283-1
26	44.8	2.0	1288	1	US-08-440-856A-9
27	44.6	2.0	320	4	US-09-165-264-11

C 28	44.6	2.0	13987	2	US-08-804-227C-13	Sequence 13, Appl
C 29	44.4	2.0	320	4	US-09-165-264-14	Sequence 14, Appl
C 30	44.4	2.0	2563	1	US-08-608-241-5	Sequence 5, Appl1
C 31	44.4	2.0	2563	2	US-08-922-182-5	Sequence 5, Appl1
C 32	44.4	2.0	2563	2	US-08-919-953-5	Sequence 5, Appl1
C 33	44.2	2.0	318	4	US-09-165-264-12	Sequence 12, Appl
C 34	44.2	2.0	320	4	US-09-165-264-13	Sequence 13, Appl
C 35	44.2	2.0	1908	1	US-08-173-508-1	Sequence 1, Appl1
C 36	44.2	2.0	1908	2	US-08-265-310-1	Sequence 1, Appl1
C 37	44.2	2.0	1908	3	US-08-951-742-1	Sequence 1, Appl1
C 38	43.6	2.0	1500	4	US-09-593-711A-10	Sequence 10, Appl
C 39	43.6	2.0	2418	1	US-08-462-484-1	Sequence 1, Appl1
C 40	43.6	2.0	2418	1	US-08-441-147-1	Sequence 1, Appl1
C 41	43.6	2.0	2418	5	PCT-US95-07536-1	Sequence 1, Appl1
C 42	43.6	2.0	3624	1	US-07-951-715A-6	Sequence 6, Appl1
C 43	43.6	2.0	3624	2	US-08-459-448A-6	Sequence 6, Appl1
C 44	43.6	2.0	3624	3	US-08-459-595A-6	Sequence 6, Appl1
C 45	43.6	2.0	3624	3	US-08-459-504B-6	Sequence 6, Appl1

ALIGNMENTS

RESULT 1  
US-08-932-376A-3  
; Sequence 3, Application US/08932376A  
; Patent No. 5869309  
; GENERAL INFORMATION:  
; APPLICANT: Politino, Michael  
; APPLICANT: Tonzi, Sean M.  
; APPLICANT: Usher, John J.  
; APPLICANT: Burnett K, William V.  
; APPLICANT: Romancik, Guna  
; TITLE OF INVENTION: CEPHALOSPORIN ESTERASE GENE FROM  
; TITLE OF INVENTION: RHODOSPORIDIUM TORULOIDES  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Bristol-Myers Squibb Company  
; STREET: Rt. 206 & Provinceline Road  
; CITY: Princeton  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 08543-4000  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/932,376A  
; FILING DATE: 17-SEP-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Savitsky, Thomas R.  
; REGISTRATION NUMBER: 31,661  
; REFERENCE/DOCKET NUMBER: ON0144a  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (609) 252-4956  
; TELEFAX: (609) 252-4526  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2220 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; US-08-932-376A-3

Query Match 100.0%; Score 2220; DB 2; Length 2220;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;











[illegible]



TELEPHONE: 312 707-9889  
TELEFAX: 312 707-9155  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1505 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Brassica napus  
STRAIN: Westar  
DEVELOPMENTAL STAGE: Somatic  
TISSUE TYPE: Root  
US-07-915-246-1

Query Match 2.38; Score 51; DB 1; Length 1505;  
Best Local Similarity 45.6%; Pred. No. 0.0054;  
Matches 221; Conservative 0; Mismatches 260; Indels 4; Gaps 1;

Qy 1442 ttcattatcactgaagccactattcagaacagacagatcgaagcaccagtcgacgagcgtc 1501  
Db 1166 tccacctgcacggccacgtccatattccctccacagaccagcgttcctccaccaccgta 1107  
Qy 1502 tccagttcagccgctcctcgcggtcttccctcacatcacctcgagagagagcgccag 1561  
Db 1106 tccaccacgggtgacacctccacctctccacggcctccatttcaccgctccaccacc 1047  
Qy 1562 gccgtcgcgaagcagtaaccgatctccagcgcgcgtccaaaggggaacacacattctctgc 1621  
Db 1046 tccatgtccacctgcacggccacgtcccatatctctccacagaccagcactctccacc 987  
Qy 1622 atctcggcgtcagtcgagacgtcgtgctgcttccctccagcgtcgtctctccagagta 1681  
Db 986 accgtatccaccacc-----atggcagaccgacgtctccaccaccacgtctccggtc 931  
Qy 1682 ttcgctgaactccgcttgcgcgagctgcgcgacctgaagcgcgaggggttcgg 1741  
Db 930 ctccacctccaccatgtctccagctctcctccgctccgtatccagctccaccaccttcac 871  
Qy 1742 ctgctcgcgcccaagggctctcttgatcagcgcgcgtcaccagcgcgacccaactc 1801  
Db 870 cgtttccaccctccataaaccagatcctctccagcagaccaccaccacgtccaccagcac 811  
Qy 1802 gactacatcgctccatctggaagcgggaagaagtcggtctcgtccgtccagtccttoga 1861  
Db 810 caccaccatgtccaccgctcctctccaccatgctccagctccaccgtaaccagcac 751  
Qy 1862 cggcgcgtcggcgtctcagagcttcaaccgcgaacacacagctgcgaacaagac 1921  
Db 750 cactctcttccgcccaccctccacacacacacacagcagatccactccataccatpacc 691  
Qy 1922 catca 1926  
Db 690 CAGCA 686

RESULT 10  
US-07-642-734C-3/c  
Sequence 3, Application US/07642734C  
Patent No. 5824513  
GENERAL INFORMATION:  
APPLICANT: Katz, L  
APPLICANT: Donadio, S  
APPLICANT: Mcalpine, J B  
TITLE OF INVENTION: Recombinant DNA Method for Producing  
Erythromycin Analogs  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Edward H. Gorman

STREET: Abbott Laboratories D377/AP6D-2 One Abbott  
STREET: Park Rd  
CITY: Abbott Park  
STATE: IL  
COUNTRY: US  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/642,734C  
FILING DATE: 17-JAN-91  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Dauckers, Andreas M  
REGISTRATION NUMBER: 32652  
REFERENCE/DOCKET NUMBER: 4952.US.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 708-937-9396  
TELEFAX: 708-938-2623  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20235 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Saccharopolyspora erythraea  
STRAIN: NRRL 238  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 19..10722  
OTHER INFORMATION: /codon\_start= 19  
OTHER INFORMATION: /function= "gene eryA"  
OTHER INFORMATION: /product= "eryA ORF2 encoding modules 3 & 4 for  
OTHER INFORMATION: 6-deoxyerythronolide B"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 19..4470  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: module 3"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 97..1482  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoacyl ACP synthase of module 3"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1693..2670  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyltransferase domain module 3"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 3406..3921  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoreductase domain of module 3"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 4171..4428  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyl carrier domain of module 3"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 4471..10722  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: module 4"  
FEATURE:  
NAME/KEY: misc\_feature

```
LOCATION: 4471..5847
OTHER INFORMATION: /function= "approximate span of
FEATURE: OTHER INFORMATION: beta-ketoacylACPsynthase domain of module 6"
NAME/KEY: misc_feature
LOCATION: 6054..7036
OTHER INFORMATION: /function= "approximate span of
FEATURE: OTHER INFORMATION: acyltransferase domain of module 4"
NAME/KEY: misc_feature
LOCATION: 7165..9216
OTHER INFORMATION: /function= "approximate span of
FEATURE: OTHER INFORMATION: dehydratase and enoylreductase domains m"
NAME/KEY: misc_feature
LOCATION: 9433..9984
OTHER INFORMATION: /function= "approximate span
FEATURE: OTHER INFORMATION: beta-ketoreductase of module 4"
NAME/KEY: misc_feature
LOCATION: 10225..10483
OTHER INFORMATION: /function= "approximate span of
FEATURE: OTHER INFORMATION: acyl carrier domain of module 4"
NAME/KEY: CDS
LOCATION: 10723..20235
OTHER INFORMATION: /codon_start= 10723
OTHER INFORMATION: /function= "gene -eryA"
OTHER INFORMATION: /product= "orf3 encoding modules 5 & 6"
FEATURE: OTHER INFORMATION: 6-deoxyerythronolide B formatio"
NAME/KEY: misc_feature
LOCATION: 10723..15165
OTHER INFORMATION: /function= "approximate span of
FEATURE: OTHER INFORMATION: module 5"
NAME/KEY: misc_feature
LOCATION: 10831..12174
OTHER INFORMATION: /function= "approximate span of
FEATURE: OTHER INFORMATION: beta-ketoacylACPsynthase domain of modul"
NAME/KEY: misc_feature
LOCATION: 12379..13350
OTHER INFORMATION: /function= "approximatr span of
FEATURE: OTHER INFORMATION: acyltransferase domain of module 5"
NAME/KEY: misc_feature
LOCATION: 14062..14610
OTHER INFORMATION: /function= "approximate span of
FEATURE: OTHER INFORMATION: beta-ketoreductase of module 5"
NAME/KEY: misc_feature
LOCATION: 14857..15114
OTHER INFORMATION: /function= "approximate span of
FEATURE: OTHER INFORMATION: acyl carrier domain of module 5"
NAME/KEY: misc_feature
LOCATION: 15166..20235
OTHER INFORMATION: /function= "approximate span of
FEATURE: OTHER INFORMATION: module 6"
NAME/KEY: misc_feature
LOCATION: 15172..16569
OTHER INFORMATION: /function= "approximate span of
FEATURE: OTHER INFORMATION: beta-ketoacylACPsynthase domain of modul"
NAME/KEY: misc_feature
LOCATION: 16768..17721
OTHER INFORMATION: /function= "approximate span of
FEATURE: OTHER INFORMATION: acyltransferase domain of module 6"
NAME/KEY: misc_feature
LOCATION: 18379..18921
```

```
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19149..19398
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19492..20235
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: thioesterase domain of module 6"
FEATURE:
US-07-642-734C-3

Query Match 2.2%; Score 49.2; DB 1; Length 20235;
Best Local Similarity 45.4%; Pred. No. 0.032;
Matches 216; Conservative 0; Mismatches 258; Indels 2; Gaps 1;

Qy 1540 catcacctcgaggagcgccggtcgcggaagcagtcacccgatctccgacgcgcgtc 1599
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17256 CACGACCGAGGAGGGGAGTTGACCGCGGCGAGGATCGGTCTCCACCGGTGGAT 17197
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1600 aaagggcaacaccttctcgcatctcgggcgtcgcgcgactcgcgactcgtgtgcgt 1659
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17196 CAGCGCGGGGCGGTTTCACCGGGAGCGGCCAACGAGACCATGCCGCCCTTGC 17137
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1660 tccccg--tggtttctcgagatttcggtgacttcccgcttgcgcgacgcgcgac 1717
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17136 CGCAGCGAACGCGCTTCGCGCGCAGGGCCACGCGCACGCCGCTCTCCAGCGAACAC 17077
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1718 ctactggacgcgcgaggttcggttcgctcgcccaagggcctcttcgactacgcgc 1777
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17076 CCGCGCCACACCGCGCGCGGATCTGCCCTGCGATGCGCGATGACGCGCGAGGGCT 17017
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1778 ggtcaccacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1837
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17016 GACTCCGCAAGCGCCCGCACAGCGCGCGCACGACACCATCAGGAGAACACACCGCTG 16957
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1838 ggtctcgtcgtccagtccttcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1897
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 16956 TAGACGCTGACCGCGCTCCAGCGACGCGCGCTCGGACGCTGCTCCAGACATTCGGAGGC 16897
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1898 gaacaacaagcgtgccacaagaccatcaacccttactggcgcgcgcgcgcgcgcgcgc 1957
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 16896 CGAGAACCGCGCGCACCTCCGACACACCGCATCGCATCGCGGATCGACTCGGCGAGAC 16837
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1958 gcagctccttcaacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2013
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 16836 GGGGACCGAGAGCAAGCCCGGGCGCATGCCCTCCACTGAGCACCCCTGACCTGGGA 16781
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
US-08-439-009A-3/c
; Sequence 3, Application US/08439009A
; Patent No. 6004787
; GENERAL INFORMATION:
; APPLICANT: Donadio, S
; APPLICANT: Katz, L
; APPLICANT: Mcalpine, J B
; TITLE OF INVENTION: Method of Directing Biosynthesis of
; TITLE OF INVENTION: Specific Polyketides
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Steven F. Weinstein
; STREET: Abbott Laboratories D377/AP6D-2 One Abbott
; STREET: Park Rd
; CITY: Abbott Park
; STATE: IL
; COUNTRY: US
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/439,009A
FILING DATE: 11-MAY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Casuto, Dianne
REGISTRATION NUMBER: 40,943
REFERENCE/DOCKET NUMBER: 4952.US.D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 847-938-3137
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 20235 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Saccharopolyspora erythraea
STRAIN: NRRL 238
FEATURE:
NAME/KEY: CDS
LOCATION: 19..10722
OTHER INFORMATION: /codon_start= 19
OTHER INFORMATION: /function= "gene eryA"
OTHER INFORMATION: /product= "eryA ORF2 encoding modules 3 & 4 for
OTHER INFORMATION: 6-deoxyerythronolide B"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19..4470
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 97..1482
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacyl ACP synthase of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1693..2670
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3406..3921
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4171..4428
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4471..10722
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4471..5847
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of module"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 6054..7026
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7165..9216
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: dehydratase and enoylreductase domains m"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 9433..9984
OTHER INFORMATION: /function= "approximate span
OTHER INFORMATION: beta-ketoreductase of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10225..10483
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 4"
FEATURE:
NAME/KEY: CDS
LOCATION: 10723..20235
OTHER INFORMATION: /codon_start= 10723
OTHER INFORMATION: /function= "gene -eryA"
OTHER INFORMATION: /product= "orf3 encoding modules 5 & 6
OTHER INFORMATION: 6-deoxyerythronolide B formatio"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10723..15165
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10831..12174
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 12379..13350
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14062..14610
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14857..15114
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15166..20235
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15172..16569
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 16768..17721
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 18373..18921
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19149..19398
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 6"
FEATURE:
NAME/KEY: misc_feature
```

```

; FILING DATE: 30-APR-1993
; APPLICATION NUMBER: JAPAN 4-126511
; FILING DATE: 19-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley, Jr., James F.
; REGISTRATION NUMBER: 27794
; REFERENCE/DOCKET NUMBER: SHGN-7
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)596-9000
; TELEFAX: (212)596-9090
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2064
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Streptomyces fradiae
; FEATURE:
; NAME/KEY: -35 signal
; LOCATION: 359..364
; IDENTIFICATION METHOD: by experiment
; NAME/KEY: -10 signal
; LOCATION: 378..383
; IDENTIFICATION METHOD: by experiment
; NAME/KEY: CDS
; LOCATION: 435..1505
; IDENTIFICATION METHOD: by experiment
; NAME/KEY: sig peptide
; LOCATION: 435..944
; IDENTIFICATION METHOD: by experiment
;
; US-08-343-428-1
;
; Query Match 2.2%; Score 49; DB 1; Length 2064;
; Best Local Similarity 44.1%; Pred. No. 0.017;
; Matches 249; Conservative 0; Mismatches 315; Indels 1; Gaps
;
QY 1513 ccgcctccctcgcgccctctccctacatcacctcgaggagcgccagggcgctcgcgaa 1572
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 875 CCAGCAGGTGCGCGCTCCCTCGAGGCGCCGTCAGCGTACC CGCGTCCCCGGGGTGT 934
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1573 gcagtagccgatctcgacgcgcgtcgaaggcgcaacaccttctctcgcatctcgccgt 1632
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 935 CCAGCGGAGGTGGCGCGCGGACGCCATCTACGGCGCGCTCGCGCTGCTCGCGCGC 994
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1633 catcgggactcgacctctggtgtgcttcccggtcggttcttcgagtgattccgctgact 1692
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 995 CTTCAACGTCACCAAGAAGC-GCGTTCGGTACTTCTTACGCGCGCGGCACTGCACCAAC 1053
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1693 tcccgcttcgcccgaactccgacactagccgcgagcgagcgcttcggtcgctccgccc 1752
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1054 TCTGTCCACCTGTGTCGTCCACTTCGGCGGCACGTTCATCGCGTCCGCGAGGACACA 1113
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1753 acaaggcgctcttcgactacgcgcgcgtccacacgcgacgcagacaactctgactacatcg 1812
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1114 GCTTCCGACCAAGACTACGGCATCGTCCGCTACAGCACACGACCAACGTGGAGGCC 1173
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1813 gctccatctggaaaggcaagtggtctctgcgtccagtcctctcagcggcgcgctcg 1872
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1174 GGGTCAACCTGTAAACGGCGGCTACAGAGATCGCCTCGCGCGGCGACGCCCTCGCTGG 1233
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1873 gcggcttcacgagacttcaaccgaaacaaacgcgtgcacaagaccatcaacctt 1932
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1234 GCAGGCCATCAAGAGAGCGGCTCCACGACCAGAGTACACAGCGGACCGCTACGCGCG 1293
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1933 actggccgacttcgactcgggcaagcagctctctcttcaaacacgacgacgagggacacc 1992
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1294 TCAACGTACCGTCACTACAGGAGCGCCCGCTCTACGGCATGTCTCCGACGACCGCT 1353
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1993 tctctcccgccaccgcgcgtctgttgagacttcaagcttgaccgactttggcagcgacc 2052
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



```

; GENERAL INFORMATION:
;
; APPLICANT: Cardineau, Guy A.
; APPLICANT: Stelman, Steven J.
; APPLICANT: Narva, Kenneth E.
;
; TITLE OF INVENTION: Plant-Optimized Genes Encoding Pesticidal Toxins
;
; FILE REFERENCE: MA-714XC2
;
; CURRENT APPLICATION NUMBER: US/09/178,252
;
; CURRENT FILING DATE: 1998-10-23
;
; EARLIER APPLICATION NUMBER: 60/065,215
;
; EARLIER FILING DATE: 1997-11-12
;
; EARLIER APPLICATION NUMBER: 60/076,445
;
; EARLIER FILING DATE: 1998-03-02
;
; NUMBER OF SEQ ID NOS: 27
;
; SOFTWARE: PatentIn Ver. 2.0

```

```

RESULT 14
US-09-130-114-2
; Sequence 2, Application US/09130114
; Patent No. 5976807
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert A.
; APPLICANT: Dama, Bassam B.
; APPLICANT: Robbins, Alan K.
; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
; TITLE OF INVENTION: From Multiple Transfected Episomes
; FILE REFERENCE: 0867/ID903US1
; CURRENT APPLICATION NUMBER: US/09/130.114
; CURRENT FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1931
; TYPE: DNA
; ORGANISM: EBNA
US-09-130-114-2

Query Match 2.2%; Score 48.2; DB 2; Length 1931;
Best Local Similarity 45.4%; Pred.No.0.026;
Matches 173; Conservative 0; Mismatches 208; Indels 0; Gaps

QY 1479 tcagcagcagtcgagcgctccacgttcgacgcctctctcgacgcctctctcgcggtctctccct 1538
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 714 tcctcgctctctctctctctctctctctctctctctctctctctctctctctctctctctctc 773

QY 1539 acatcacctcggaggagcgccggcgctcggaagcagatcacccgatctcggacgcgcgt 1598
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 774 tctcctctctctctctctctctctctctctctctctctctctctctctctctctctctctc 833

QY 1599 caaagggaacacctctctctctctctctctctctctctctctctctctctctctctctctc 1658
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 834 tctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctc 893

QY 1659 ttccctcgctctctctctctctctctctctctctctctctctctctctctctctctctctctc 1718
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 894 tccctctctctctctctctctctctctctctctctctctctctctctctctctctctctctc 953

QY 1719 tactgagccgcgagcgctctctctctctctctctctctctctctctctctctctctctctc 1778
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 954 tctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctc 1013

QY 1779 gctcacgcgcgacgagcaactctgactacatcggtctctctctctctctctctctctctctc 1838
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1014 agtctctctctctctctctctctctctctctctctctctctctctctctctctctctctctc 1073

QY 1839 gtctctctctctctctctctctctctctctctctctctctctctctctctctctctctctc 1859
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1074 acttctctctctctctctctctctctctctctctctctctctctctctctctctctctctc 1094

RESULT 15
US-09-178-252-26
; Sequence 26, Application US/09178252
; Patent No. 6218188
; GENERAL INFORMATION:
; APPLICANT: Cardineau, Guy A.
; APPLICANT: Stelman, Steven J.
; APPLICANT: Narva, Kenneth E.
; TITLE OF INVENTION: Plant-Optimized Genes Encoding Pesticidal Toxins
; FILE REFERENCE: MA-714XC2
; CURRENT APPLICATION NUMBER: US/09/178.252
; CURRENT FILING DATE: 1998-10-23
; EARLIER APPLICATION NUMBER: 60/065,215
; EARLIER FILING DATE: 1997-11-12
; EARLIER APPLICATION NUMBER: 60/076,445
; EARLIER FILING DATE: 1998-03-02
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.0

```

Search completed: June 29, 2002, 15:07:14  
Job time: 20670 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 29, 2002, 13:05:40 ; Search time 4149.72 seconds  
(without alignments)  
7220.545 Million cell updates/sec

Title: US-09-801-852A-3  
Perfect score: 2220  
Sequence: 1 ggatcacccgaactctgtc.....agacgttgcaagtgcgaaa 2220

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

#### Database :

EST.\*  
1: em\_estba.\*  
2: em\_esthum.\*  
3: em\_estin.\*  
4: em\_estma.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_hic.\*  
9: gb\_estl.\*  
10: gb\_est2.\*  
11: gb\_hic.\*  
12: gb\_gss.\*  
13: em\_gss\_hum.\*  
14: em\_gss\_inv.\*  
15: em\_gss\_pln.\*  
16: em\_gss\_vrt.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
C 1	64.2	2.9	828	12 CNS04GHZ	AL289664 Tetraodon
C 2	60.6	2.7	793	12 CNS01T34	AL166009 Tetraodon
C 3	59.8	2.7	1101	12 CNS016MU	AL106944 Drosophila
C 4	59.4	2.7	1101	12 CNS00GZV	AL072957 Drosophila
C 5	59.2	2.7	625	9 AA202160	AA202160 LD02542.5
C 6	59.2	2.7	649	9 AA246278	AA246278 LD05065.5
C 7	59.2	2.7	673	10 B1574042	B1574042 RH09457.5
C 8	59.2	2.7	676	10 B1564550	B1564550 RH61852.5
C 9	58.6	2.6	880	12 CNS04463	AL273684 Tetraodon
C 10	58.4	2.6	925	12 CNS0091P	AL053013 Drosophila
C 11	58.2	2.6	566	12 CNS03JN0	AL247077 Tetraodon
C 12	58.2	2.6	1463	10 BM320858	BM320858 rumast-1
C 13	58	2.6	1464	3 BE636746	BE636746 rockefell
C 14	57.6	2.6	1011	12 AG126307	AG126307 Pan trogl
C 15	57.2	2.6	568	9 AI516721	AI516721 LD42723.5
C 16	57.2	2.6	855	9 AL572700	AL572700 AL572700
C 17	57.2	2.6	873	12 AG136937	AG136937 Pan trogl

C 18	57	2.6	631	10 B1955992	B1955992 HVSMEm002
C 19	57	2.6	917	12 AG081214	AG081214 Pan trogl
C 20	56.8	2.6	857	12 AG044313	AG044313 Pan trogl
C 21	56.4	2.5	586	10 B1626691	B1626691 RH67704.5
C 22	56.4	2.5	703	12 AQ738630	AQ738630 HS_5381_LB
C 23	56.4	2.5	1101	12 CNS0153F	AL104949 Drosophila
C 24	56.2	2.5	829	9 AW448067	AW448067 BRY_1587
C 25	56.2	2.5	988	12 AG135153	AG135153 Pan trogl
C 26	56.2	2.5	1321	12 AG126084	AG126084 Pan trogl
C 27	56	2.5	1037	12 AG146864	AG146864 Pan trogl
C 28	56	2.5	1101	12 CNS001T4	AL061330 Drosophila
C 29	55.8	2.5	781	12 AG153560	AG153560 Pan trogl
C 30	55.6	2.5	1223	10 B1416573	B1416573 hasp001xg
C 31	55.4	2.5	867	10 B1958202	B1958202 HVSMEm001
C 32	55.4	2.5	1064	12 AG134261	AG134261 Pan trogl
C 33	55.4	2.5	1083	10 BM460150	BM460150 AGENCOURT
C 34	55.4	2.5	1101	12 CNS01523	AL104901 Drosophila
C 35	55.2	2.5	768	12 AG138112	AG138112 Pan trogl
C 36	55.2	2.5	932	12 AG134650	AG134650 Pan trogl
C 37	55.2	2.5	1052	10 B1416480	B1416480 hasp001xa
C 38	55.2	2.5	1082	10 BE455154	BE455154 HVSMEm009
C 39	55.2	2.5	1155	12 AG124478	AG124478 Pan trogl
C 40	55	2.5	820	10 BG320478	BG320478 zm03_09f0
C 41	55	2.5	976	10 B1416536	B1416536 hasp001xe
C 42	54.8	2.5	809	10 B1952732	B1952732 HVSMEm000
C 43	54.8	2.5	878	12 AZ183970	AZ183970 SP_1002.A
C 44	54.8	2.5	926	12 CNS03C4D	AL237334 Tetraodon
C 45	54.8	2.5	933	12 AG126371	AG126371 Pan trogl

#### ALIGNMENTS

RESULT 1  
CNS04GHZ/c  
LOCUS  
DEFINITION  
CNS04GHZ 828 bp DNA linear GSS 21-MAY-2000  
Tetraodon nigroviridis genome survey sequence PUC-ori end of clone  
108D17 of library G from Tetraodon nigroviridis, genomic survey  
sequence.  
AL289664  
AL289664.1 GI:8028243  
GSS: genome survey sequence.  
KEYWORDS  
SOURCE  
Tetraodon nigroviridis.  
ORGANISM  
Tetraodon nigroviridis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Perciformes; Tetraodontiformes;  
Tetraodontidae; Tetraodon.  
REFERENCE  
1 (bases 1 to 828)  
Roest-Crollius H., Jaillon O., Dasilva C., Fizes C., Fisher C.,  
Bouneau L., Billault A., Quetier F., Saurin W., Bernot A. and  
Weissenbach J.  
Characterization and repeat analysis of the compact genome of the  
freshwater pufferfish Tetraodon nigroviridis  
Unpublished  
2 (bases 1 to 828)  
Roest-Crollius H., Jaillon O., Dasilva C., Bouneau L., Fisher C.,  
Bernot A., Fizes C., Wincker P., Brottier P., Quetier F.,  
Saurin W. and Weissenbach J.  
Human gene number estimate provided by genome wide analysis using  
Tetraodon nigroviridis DNA sequence  
Unpublished  
3 (bases 1 to 828)  
Genoscope.  
Direct Submission  
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases  
This sequence is a single read and was generated as part of a large  
scale clone-end sequencing project of the Tetraodon nigroviridis  
genome. For more information, please take a look at  
http://www.genoscope.cns.fr/tetraodon.  
Location/Qualifiers  
1. 828  
/organism="Tetraodon nigroviridis"





```
Query Match
Best Local Similarity 2.7%; Score 59.2; DB 9; Length 625;
Matches 122; Conservative 0; Mismatches 78; Indels 6; Gaps 1;

QY 434 tcggaggattgctcttctcaatgctgtgccccgcggctctgtgagggcgacaat 493
DB 305 TCCGAGGACTGCTTTCTTCTTATGTATATGACCTAAAGTCAAAAGTACGGGACTCCT 364

QY 494 cttccgctctctctacacagagtgctcagccttcgagtgatcgagcaccgac 553
DB 365 TTACCGGTGATGCTGTGGATTTCATGTGGAGGCTTCTTCTCGCAACGGCAACAGTGAC 424

QY 554 agcgacttcgcttcaccagcacagcggaacaaagatggtcgtttaaattctccag 613
DB 425 TTCCACTTTCCCG-----CCAAAGCTCATGAGGAGGAGTTCATGTGTGTCACCTGTAAT 478

QY 614 taccgtctcggcagctttggtttct 639
DB 479 TACCAGACTTGGAGCGCTTGGTTTCT 504

RESULT 6
LOCUS AA246278 649 bp mRNA linear EST 19-APR-2001
DEFINITION LD05065.Sprime LD Drosophila melanogaster embryo Bluescript
Drosophila melanogaster cDNA clone LD05065 Spime, mRNA sequence.
ACCESSION AA246278
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 649)
AUTHORS Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
Lewis,S. and Rubin,G.M.
BDGP/HMI Drosophila EST Project
Unpublished (2001)
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: 50 row: F column: 5
High quality sequence stop: 539.
FEATURES
Location/Qualifiers
1..649
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="LD05065"
/clone.lib="LD Drosophila melanogaster embryo Bluescript"
/sex="male and female"
/dev_stage="0 to 24 hours mixed stage embryonic"
/lab_host="SOUR"
/note="Organ: embryo; Vector: Bluescript SK; Site:1: EcoRI
; Site:2: XhoI; Constructed using Stratagene ZAP-cDNA
Synthesis kit. Oligo dt-primed and directionally cloned at
EcoRI and XhoI in Bluescript SK(+/-)"
BASE COUNT 153 a 150 c 184 g 161 t 1 others
ORIGIN

Query Match
Best Local Similarity 2.7%; Score 59.2; DB 9; Length 649;
Matches 122; Conservative 0; Mismatches 78; Indels 6; Gaps 1;

QY 434 tcggaggattgctcttctcaatgctgtgccccgcggctctgtgagggcgacaat 493
DB 351 TCCGAGGACTGCTTTCTTCTTATGTATATGACCTAAAGTCAAAAGTACGGGACTCCT 410

QY 494 cttccgctctctctacacagagtggtggtcagccttcgagtgatcgagcaccgac 553
```

```
DB 411 TTACCGGTGATGCTGTGGATTTCATGTGGAGGCTTCTTCTCGCAACGGCAACAGTGAC 470
QY 554 agcgacttcgcttcaccagcacagcggaacaaagatggtcgtttaaattctccag 613
DB 471 TTCCACTTTCCCG-----CCAAAGCTCATGAGGAGGAGTTCATGTGTGTCACCTGTAAT 524
QY 614 taccgtctcggcagctttggtttct 639
DB 525 TACCAGACTTGGAGCGCTTGGTTTCT 550

RESULT 7
LOCUS BI574042 673 bp mRNA linear EST 06-SEP-2001
DEFINITION RH09457.Sprime RH Drosophila melanogaster normalized Head pflc-1
Drosophila melanogaster cDNA clone RH09457 5 similar to clt;
FBAN000958 GO: [carboxyesterase (GO:0004091); carboxyesterase
(GO:0004091)] located on: 2R 57F4-57F4;: 07/26/2001, mRNA sequence.
ACCESSION BI574042
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 673)
AUTHORS Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B., Carlson
,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,B., George
,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Misra,S.,
Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and Rubin
,G.M.
BDGP/HMI RH Drosophila EST Project
Unpublished (2001)
Other ESTs: RH09457.3prime
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
hit genomic AE003454: arm:2R [16312727,16613591]
estimated-cyto:57D11-57F8: 07/26/2001
Plate: RH.94 row: E column: 9
High quality sequence stop: 577.
FEATURES
Location/Qualifiers
1..673
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="RH09457"
/clone.lib="RH Drosophila melanogaster normalized Head
pflc-1"
/sex="male and female"
/dev_stage="Adult"
/lab_host="DH5-alpha Tona"
/note="Organ: head; Vector: pFlc1; Site:1: XhoI; Site:2:
BamHI; Library was kindly generated by Piero Carninci at
the RIKEN. The library was normalized and excised using
Cre recombinase. Plasmid cDNA library."
BASE COUNT 157 a 156 c 191 g 168 t 1 others
ORIGIN

Query Match
Best Local Similarity 2.7%; Score 59.2; DB 10; Length 673;
Matches 122; Conservative 0; Mismatches 78; Indels 6; Gaps 1;

QY 434 tcggaggattgctcttctcaatgctgtgccccgcggctctgtgagggcgacaat 493
DB 374 TCCGAGGACTGCTTTCTTCTTATGTATATGACCTAAAGTCAAAAGTACGGGACTCCT 433
```

[illegible]

RESULT	8
BIS64550	
LOCUS	BIS64550
DEFINITION	RH61852.1 prime RH Drosophila melanogaster normalized Head Pflc-1
LOCUS	Drosophila melanogaster cDNA clone RH61852.5 similar to clt;
ACCESSION	F8an0009858 GO:[carboxyesterase (GO:0004031); carboxyesterase (GO:0004091)] located on: 2R 57F4-57F4:: 08/23/2001, mRNA sequence.
VERSION	BIS64550.1 GI:15453742
EST	EST 06-SEP-2001

KEYWORDS	EST.
SOURCE	fruit fly.
ORGANISM	<i>Drosophila melanogaster</i>
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 676)
AUTHORS	Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B., Carlson, J., Champagne, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G., Misa, S., Mungall, C. J., Nuno, J., Pacleb, J., Paragas, V., Park, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S. E., Celis, P. K., and Rubin

TITLE BGP/HHMI RH Drosophila EST Project  
JOURNAL Unpublished (2001)  
COMMENT Contact: Stapleton, M.  
BDGP  
Lawrence Berkeley National Lab  
One Cyclotron Rd, Berkeley, CA 94720, USA  
Fax: 510 486 6798  
Email: [http://www.fruitfly.org/EST\\_est@fruitfly.berkeley.edu](http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu)  
hit genomic AE003454: arm:2R [16312727..16613591]  
estimated-cyto:57D11-57F8: 08/23/2001  
Plate: RH.618 row: E column: 4  
High quality sequence stop: 579.

```

FEATURES
source
Location/Qualifiers
1. .676
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="RH61852"
/clone_lib="RH Drosophila melanogaster normalized Head
pPic-1"
/sex="male and female"
/dev_stage="Adult"
/lab_host="DH5-alpha TONa"
/note="Organ: head; Vector: pPic1; site.1: xhoI; site.2:
BamHI; Library: was kindly generated by Piero Carninci at
the RIKEN. The library was normalized and excised using
cre recombinase. Plasmid cDNA library."
157 a 159 c 191 q 168 t 1 others
BASE COUNT

```

BASE COUNT 157 a 159 c 191 g 168 t 1 others  
ORIGIN

Query Match 2.7% Score 59.2; DB 10; Length 676;  
Best Local Similarity 59.2%; pred. No. 1.2;  
Matches 122; Conservative 0; Mismatches 78; Indels 6; Gaps 1;

Qy 434 tcggagattgctcttctctaattgtcgttgccccccgcgcgtcgctgtagaggcgacaat 493

D8 374 TCCGAGACTGCCTTTTTCCTTAATGTATACACCTAAAGTCAAAAGTACGGGACTCCT 433

Qy	494	cttcgcgctcgtctctacattcaacggagggtggtctacgcccttcgcgcgagtgcgagcaacggc	553
Db	434	TTACCCGTGATGGTCTGGATTATGGTGGAGGCTTCTTTCGGCAACGGCAACAGTGAC	493
Qy	554	aggaactttgccgecttcaaccaagcacacgggaaccaagatggctcgtgttaaatctccag	613
Db	494	TTCCACTTCCCG-----CCAAGCTCATGAGCAGGAGGTTCATTGTGGTCACCCGTAAT	547
Qy	614	taccgctctcggcagctttgtttctct	639
Db	548	TACCGACTTGGAGCGCTTGTTTCTCT	573

RESULT	9
CNS04463/C	
LOCUS	
DEFINITION	CNS04463 880 bp DNA linear GSS 18-MAY-2000 Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone 080P10 of library G from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION	AL273684
VERSION	AL273684.1 GI:7995946
KEYWORDS	GSS; genome survey sequence.
SOURCE	Tetraodon nigroviridis.
ORGANISM	Tetraodon nigroviridis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.  
1 (bases 1 to 880)

Roest-Crolius H., Jaillon O., Dasilva C., Fizames C., Fisher C., Bounseul J., Billault A., Quetier F., Saurin W., Bernot A. and Weissenbach J.

Characterization and repeat analysis of the compact genome of the freshwater pufferfish *Tetraodon nigroviridis*

JOURNAL REFERENCE	2 (bases 1 to 880)
AUTHORS	Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Barnot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.
TITLE	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL REFERENCE	Unpublished 3 (bases 1 to 880)
AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL	Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <a href="http://www.genoscope.cns.fr/Tetraodon">http://www.genoscope.cns.fr/Tetraodon</a> .

```

FEATURES
source
Location/Qualifiers
1. .880
/organism="Tetradodon nigroviridis"
/db_xref="taxon:99883"
/clone="080P10"
/clone_lib="G"
/note="Genoscope sequence ID : C0BG080DH05S1-end :
plc-O-1"

```

BASE COUNT	87 a	96 c	518 g	109 t	70 others
ORIGIN					
Query Match		2.6%	Score 58.6;	dB 12;	Length 880;
Best Local Similarity		44.0%;	Pred. No. 1.6;		
Matches 157;		Conservative	18;	Mismatches 182;	Indels 0;
Gaps 0;					

Qy	5	ccaccgaaactctgcgcgctttcttggtttcttctctgtgctgcgcacgcgccttcc	64
Db	442	ccc	383
Qy	65	cgactgcgcgcacatgctcttaacctcttacctcgcctcctcgcctgcgcacgtccag	124
Db	382	ccc	323

[illegible]

RESULT	10
CNS0091P/c	
LOCUS	GSS 925 bp DNA linear GSS 03-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION	AL053013
VERSION	AL053013.1 GI:4934461
KEYWORDS	GSS.
SOURCE	fruit fly.
ORGANISM	Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 925)
REFERENCE	Genoscope.
AUTHORS	Direct Submission
TITLE	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
JOURNAL	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <a href="http://www.fruitfly.org">http://www.fruitfly.org</a> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osogawa and Aron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPci-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at <a href="http://bacpac.med.buffalo.edu/drosophila_bac.htm">http://bacpac.med.buffalo.edu/drosophila_bac.htm</a> .
COMMENT	

```

FEATURES
  source
    Location/Qualifiers
      i. .925
        /organism="Drosophila melanogaster"
        /db_xref="taxon:7227"
        /clone_lib="RPCI-98"
        /clone="BACR19D16"
        /note="end : TET3"
BASE COUNT      120 a      61 c      61 g      172 t      511 others
ORIGIN

Query Match      2.6%; Score 58.4; DB 12; Length 925;
Best Local Similarity 11.5%; Pred. No. 1.7;
Matches 41; Conservative 178; Mismatches 139; Indels 0; Gaps 0;

```

QY	63	cgcgaactcggcgcatgctcctaactcttcacccctgcctccccctcgctcgagcgctcc	122
		:: :  :: ::: : :: : :  ::: : : : : : : : : : :	
Db	919	CSCCSBSCSSSMGTSSNSBCSCSSBSSTTSMWSBSBSBSGSSSSSGTSS	860
QY	123	agtcgcctttgcctctcgaactcctcctgtcgcgcgcacgaccacaagaccccctc	182

859	ACVKCNASSCGCGCGMABCCWSSSSCCGSASARGVAVRASGGAGKRGSGGSAS	800
183	cdgtctgcactctggctacgcgcgtaccaggtacttgaacgagacgcgcgactct	242
799	HSSSACBSSSSCSACWSSSSASSSSRSRGGAGSGASSRSSSSSSASAGSV	740
243	actgggtggcggaaatcgcctacgcctcgcgtcagcgtctccagggtctctcagacgcgc	302
739	VSSASSSSSCSSSVSCSVASSMSCSBSSSSASASSSSSSSSASCSCCCTSWCS	680
303	cgagcgacaaggcgtccgcgaacgcgactagtgatgacgcgactgttgcgcggtacgc	362
679	CSTSSMAARSSSSSSSCSSSMGASSASSSSSSSSSSSSSGGACBSWSGG	620
363	aggaaccaaagacaaagggttgcgcgcctagcaacgctcgagcgcgcgc	420
619	GSQSVASSCGMSVSSSGGRSGSGGGGCGVGSSSSGSGSGSGSGVCSGSGCGM	562

RESULT	11
CNS03JNO/c	
LOCUS	
DEFINITION	CNS03JNO linear GSS 17-MAY-2000 Tetraodon nigroviridis genome survey sequence pUC-Ori end of clone O31B04 of library G from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION	AL247077
VERSION	AL247077.1 GI:7968089
KEYWORDS	GSS; genome survey sequence.
SOURCE	Tetraodon nigroviridis.
ORGANISM	Tetraodon nigroviridis Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon. 1 (bases 1 to 566)
REFERENCE	Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C., Bonneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis

JOURNAL REFERENCE AUTHORS	freshwater putrefactive tetraodon nigroviridis 2 (bases 1 to 566) Roest-Crolius,H., Jalllon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Firames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.
TITLE	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL REFERENCE AUTHORS TITLE JOURNAL COMMENT	Unpublished 3 (bases 1 to 566) Genoscope. Direct Submission Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <a href="http://www.genoscope.cns.fr/Tetraodon">http://www.genoscope.cns.fr/Tetraodon</a> .

```

FEATURES
  source
    1. .566
      Location/Qualifiers
        /organism="Tetraodon nigroviridis"
        /db_xref="taxon:99883"
        /clone="031B04"
        /clone_lib="C"
        /note="Genoscope sequence ID : COBG031DA02SP1-end :
        PUC-ori"
BASE COUNT      51 a      52 c      332 g      70 t      61 others
ORIGIN

```

Query Match	2.6%	Score 58.2;	DB 12;	Length 566;
Best Local Similarity	42.6%;	Pred. No. 1.7;		
Matches 146;	Conservative	34;	Mismatches	162;
			Indels	1;
			Gaps	1;



[illegible]



FEATURES	Location/Qualifiers	BASE COUNT
source	1. 568	129 a 131 c 159 g 149 t
	/organism="Drosophila melanogaster"	
	/db_xref="taxon:7227"	
	/clone="LD4273"	
	/clone_lib="LD Drosophila melanogaster embryo pOT2"	
	/sex="male and female"	
	/dev_stage="0 to 24 hours mixed stage embryonic"	
	/lab_host="X11 Blue"	
	/note="Organ: embryo; Vector: pOT2; Site_1: EcoRI; Site_2: XhoI; Sized fractionated cDNAs were directly ligated into pOT2."	

[illegible]

Search completed: June 29, 2002, 13:05:51  
Job time: 13722 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 29, 2002, 15:03:51 ; Search time 7060.77 seconds  
(without alignments)  
6579.581 Million cell updates/sec

Title: US-09-801-852A-3  
Perfect score: 2220  
Sequence: 1 ggatccaccggaactctgtc.....agacgttgcaagtcgaaa 2220

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 1797656 seqs, 10463268293 residues  
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_v1.\*
- 15: gb\_v2.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_v1.\*
- 30: em\_v2.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htgo\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match Length	ID	Description
-----				

1	2220	100.0	2220	6	AR034145
2	2220	100.0	2220	6	AX268013
3	2220	100.0	2220	6	BD005939
4	1375	61.9	1738	6	AR034144
5	1375	61.9	1738	6	BD005938
6	1373.4	61.9	1738	6	AF025410
7	1353	60.9	1716	6	AX268011
8	74.2	3.3	298166	2	AC087563
9	73.8	3.3	165556	2	AC087190
10	73.4	3.3	303091	2	AC084799
11	71.8	3.2	220469	2	AC074307
12	69.8	3.1	181988	2	AC090552
13	68	3.1	150120	8	AP001552
14	67.8	3.1	171574	2	AC012300
15	67.6	3.0	1635	8	GC02524
16	67.4	3.0	125020	9	AF429315
17	66	3.0	1635	8	GC02622
18	66	3.0	1692	6	E02678
19	64.4	2.9	1635	6	A48376
20	64.4	2.9	1635	8	GC02387
21	64.4	2.9	1855	8	CCLIP3
22	64.2	2.9	7218	6	I66494
23	63.4	2.9	129585	8	AC090120
24	62.8	2.8	1635	8	GCTAGL
25	62.8	2.8	1635	8	GC02525
26	62.8	2.8	1635	8	GC02541
27	62.8	2.8	1635	8	GC02625
28	62.8	2.8	1828	6	A74255
29	61.4	2.8	10712	6	E32986
30	61.2	2.8	1635	8	GC02623
31	61.2	2.8	1674	6	E02497
32	61.2	2.8	1767	8	GCALIP2
33	61.2	2.8	1783	8	AB000260
34	61.2	2.8	1812	8	CCLIP5
35	59.4	2.7	265537	2	AC087228
36	59.2	2.7	1834	3	AF233526
37	59.2	2.7	1860	3	AY051959
38	59.2	2.7	29705	2	AC097965
39	59.2	2.7	150332	2	AC096099
40	58.6	2.6	102242	2	AP004043
41	58.4	2.6	78220	2	AC023212
42	58.2	2.6	45313	1	SCD95A
43	58.2	2.6	135119	2	AC011578
44	58	2.6	2043	8	CCLIP2
45	57	2.6	154409	2	AP003724

ALIGNMENTS

RESULT	1	AR034145	Sequence 3 from patent US 5869309.	2220 bp	DNA	linear	PAT 29-SEP-1999
AR034145	LOCUS	AR034145	Sequence 3 from patent US 5869309.	2220 bp	DNA	linear	PAT 29-SEP-1999
DEFINITION	AR034145	Sequence 3 from patent US 5869309.	2220 bp	DNA	linear	PAT 29-SEP-1999	
ACCESSION	AR034145	Sequence 3 from patent US 5869309.	2220 bp	DNA	linear	PAT 29-SEP-1999	
VERSION	AR034145.1	GI:5949750	2220 bp	DNA	linear	PAT 29-SEP-1999	
KEYWORDS	AR034145.1	GI:5949750	2220 bp	DNA	linear	PAT 29-SEP-1999	
SOURCE	Unknown.	Unknown.	2220 bp	DNA	linear	PAT 29-SEP-1999	
ORGANISM	Unknown.	Unknown.	2220 bp	DNA	linear	PAT 29-SEP-1999	
REFERENCE	1 (bases 1 to 2220)	Unknown.	2220 bp	DNA	linear	PAT 29-SEP-1999	
AUTHORS	Polittino, M., Tonzi, S.M., Usher, J.J., Burnett, W.V. and Romancik, G.	Unknown.	2220 bp	DNA	linear	PAT 29-SEP-1999	
TITLE	Cephalosporin esterase gene from Rhodospiridium toruloides	Unknown.	2220 bp	DNA	linear	PAT 29-SEP-1999	
JOURNAL	Patent: US 5869309-A 3 09-FEB-1999;	Unknown.	2220 bp	DNA	linear	PAT 29-SEP-1999	
FEATURES	Location/Qualifiers	Unknown.	2220 bp	DNA	linear	PAT 29-SEP-1999	
source	1..2220	Unknown.	2220 bp	DNA	linear	PAT 29-SEP-1999	
BASE COUNT	389 a 814 c 554 g 463 t	Unknown.	2220 bp	DNA	linear	PAT 29-SEP-1999	
ORIGIN	Unknown.	Unknown.	2220 bp	DNA	linear	PAT 29-SEP-1999	
Query Match	100.0%	Score 2220;	DB 6;	Length 2220;			
Best Local Similarity	100.0%	Pred. No. 0;					

Matches 2220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	ggaacacccgaactctgtcccggtttctggctttcttctctgtgtcgcccatcgct	60
Db	1	GGATCACCACCGAATCTGTGCCGCTTTTGGCTTTCTCTTGTGTGCCCCCATCGCT	60
Qy	61	ttccgactcgcgcgcactgtcttaacctcttaacctcgcctccctcgtcgcagcgt	120
Db	61	TTCCCGACTCGCGGCATGCTCTTAACCTCTTCACTCGCTCCCTCGCTCGACGCT	120
Qy	121	caagctgcgtttgctcttcgagactcctcgttcgcgcgcagcaaccccaaacagccccc	180
Db	121	CCAGCTCGCCTTTGGCTCTCCGACTCCCTCGTCGCGCGCAGCAACCAACAGAGCCCCC	180
Qy	181	tcccgtcgtcacctcggctacgcgcgtacaaaggctactttgaacgagaccgcgact	240
Db	181	TCCCGTGTGACCTCGGTAGCCCGGTACAAAGGTACTTTGAACGAGACCGCCGGACT	240
Qy	241	ctactggtggcggaaatcgcgtacgcctcggtcagcgttccaggtcctctcagacgcc	300
Db	241	CTACTGTTGGCGGGAATCCGCTACGCTCGGCTCAGCGCTTCCAGGCTCCTCAGACGCC	300
Qy	301	cgcgacgcacaaggccgctccgcaacgcgaactgagtatggaccgactctgttggccggtcgt	360
Db	301	CGCGACGCACAAGGCGCTCCGCAACGCACGTAGTATGGACCGGATCTGTTGGCCGGGTAG	360
Qy	361	cgaaggaaacaaacagaccaaagggtttgcgcgcgtacgaacagctcgcgacgcgcgc	420
Db	361	CGAGGGAACCAACACAGACCAAGGGCTTGCGCCGCCCTAGCAACAGCTCGAGCAGCGGCC	420
Qy	421	gcagaaacaggcgtcggaggaatgcctcttctcctaattcgttgcgccgcgcgcgtcgtg	480
Db	421	GCAGAACAGGCGTCGGAGGATTGCTCTTCTCAATGTCTGTCCGCCGCCGGCTCGTG	480
Qy	481	cgaaggcacaatctcccgtcctcgtctacatcagcagagtggtctacgcccttcgcgcga	540
Db	481	CGAGGGGGACAATCTTCCCGTCTCTGTTCTACATTCACGGAGTGGCTACGCCCTTCGGCGA	540
Qy	541	tgcgacacccggcagcgaactttgcgcgcttcaccaagcacacgcgggaaccagaatggtcgt	600
Db	541	TGCGAGCACCGGCAGCGACTTTTGGCGCTTCAACGAAGCACACGGGAACCAAGATGGTCGT	600
Qy	601	tgtaaatccagtaacgctctcggcagcttggtttctcgtcgtggcaagccatgaagga	660
Db	601	TGTAAATCTCAGTACCGTCTCGGACGCTTTGGTTTCTCTGCTGGCCAAAGCCATGAAGGA	660
Qy	661	ctacggtgtaaacgaacgcgcggttgcctgacacaggtgagttccgcgatgataccgcgcc	720
Db	661	CTACGGTGTAAACGACCGCGGCTTGTGTACCAAGGTGAGTTTCCGGCATGATACCCGCC	720
Qy	721	accttgcactcatgctgcgcctctccogctcgcagcaattcgccttccaatgggttca	780
Db	721	ACCTTTTGCATCATGTGTGAGCTCTCCGCTTCGCAGCAATTCGCCCTTCAATGGGTCA	780
Qy	781	acagcagctctgaattcgcgcggaaccccgatcaccttacgatttgggcagtgctgc	840
Db	781	ACAGCAGCTCTGAAGTTTGGCGGAACCCCGATCACGTTACGATTTGGGGCGAGTCTGC	840
Qy	841	aggcaggggtccgttatgaaccagatcatcgcgaacgtgagccacccgaaccgatccc	900
Db	841	AGGCGCAGGTCCTGTTATGAACAGATCATTCGGAACGTGAGCCACCGCAACCGATCTCC	900
Qy	901	agccgaactttcc	960
Db	901	AGCCGACTTTTCC	960
Qy	961	ccgtcaaggctctcgtctcaagaagccctcttccacgctgcacatggctcctccgtct	1020
Db	961	CCGTCAAGGCTCTCGGTCTCAAGAGGCCCTCTTCCAGCTGCCATCGGCTCCTCCGTCT	1020
Qy	1021	tcttccctaccagaagcaagtacaactcccccttcgcgcagctgctctactctcccaactcg	1080
Db	1021	TCTTCCCTACCAAGCAAGTACAATCCCCCTTCGCGAGCTGCTACTTCCCAACTCG	1080

[illegible]

D**b** 541 TGGGAGCACGGCAGCGACTTGGCGCTTCACCAAGCACACGGGAACCAAGATGGTCGT 600





D	b	2101	GTCCTCAGCGCTTCCCTTCCGACATTCCTTCCTTTGGTGTTATTATCTTCGAGTTC	2160
Q	y	2161	cgttgatcgccattcgtgctgtagtcactcagtagtatagcgttggaagtgcgaaa	2220
D	b	2161	CCTGTATCGGCATTGTCGCGTAGCTCACTCAGTATAGACGTTGCCAAGTCGAAA	2220
RESULT 4				
ARO34144				
LOCUS				
DEFINITION Sequence 1 from patent US 5869309.				
ACCESSION ARO34144				
VERSION ARO34144.1 GI:5949749				
KEYWORDS .				
SOURCE Unknown.				
ORGANISM Unknown.				
REFERENCE 1. (bases 1 to 1738)				
AUTHORS Polliano,M., Tonzi,S.M., Usher,J.J., Burnett,W.V. and Romancik,				
TITLE Cephalosporin esterase gene from Rhodosporidium toruloides				
JOURNAL Patent: US 5869309-A 1 09-FEB-1999;				
FEATURES Location/Qualifiers				
source 1..1738				
BASE COUNT 327 a 639 c 444 g 328 t				
ORIGIN				
Query Match 61.9%; Score 1375; DB 6; Length 1738;				
Best Local Similarity 84.7%; Pred.No.3.6e-203;				
Matches 1738; Conservative 0; Mismatches 0; Indels 313; Gaps				
Q	y	67	actgcggccatgctcctaataacctttcacacctcgctcccctcgctgcgacgtccagct	126
D	b	1	ACTCGCGCCATGCTCTTAACCTCTCACCTCGCTCCCTCGCTCGGAGCTCCAGCT	60
Q	y	127	cgcttgttcctctcagactccctctcgcgcgcgaacaccaaacgagccccctcccg	186
D	b	61	CGCCTTTGCGCTCTCCGACCTCCCTCGTCGCGCGACGAACCCAAAGAGCCCTCCCGT	120
Q	y	187	cgtcgacctcgctagcgcgcgtaccaggctacttgaacgagacgcgcgactactg	246
D	b	121	CGTCGACCTCGGTACGCCCGCTACCAGAAGCTACTTTGAACGAGACCGCGGACTCTACTG	180
Q	y	247	gtggcgcggaatccgctagcctcgctcgctcagcgcttcaggctccctcagagcccgcgac	306
D	b	181	GTGCGCGGAATCGGTACGCTCGCTCGCTCAGCGCTTCAGAGCTCTCATAGCCCGCGAC	240
Q	y	307	gcacaagcgctccgcgaacgcgactagtatggaocgatctgtgccggctagcgagg	366
D	b	241	GCACAAGCGCTCGCAACGCGACTGAGTATGGACCGATCTGTTGCCGGGTAGCGAGG	300
Q	y	367	aaccaacacagcaaaagggttgcgcgcgctagcaacagctcgacgagcgcgcgcaaa	426
D	b	301	AACCAACAGCACCAAGGGTTGCGCGCGCTAGCAACAGCTCGACGACGCGCGCGACAA	360
Q	y	427	acaggcgtcgaggatgtcctctcctaagtctgctgtcccccgcgcgctcgtgcgagg	486
D	b	361	ACAGGCGTCGAGAGATTGCCCTCTCTCAATTGCTGTGCCCCCGCGGCTCGTGGCAGGG	420
Q	y	487	cgacaattctcccgctcctgctctacattcacggagggtggtacgccttcggcgatcgag	546
D	b	421	CGACAATCTTCCCGTCTCTGCTCTACATTACGGAGGTGCTACGCTTCGCGGATGCGAG	480
Q	y	547	caccggcagcgacttgcgcgccttcaccaagcacacgcgggaacccaagatggttcgttaaa	606
D	b	481	CACCGGACGGGACTTTGCGCGCTTCACCAAGCACACGCGGAACCAAGATGGTCTGTTGTA	540
Q	y	607	tctcadtacgctcggcagcttbtgtttcctcgtgcgaagcatgaagaactacgg	666
D	b	541	TCTCCAATACCGTCTGGCAGCTTTGGTTCTCTGCTGGCCCAAGCATGAAGGACTACGG	600
Q	y	667	tgtaacacacgcgcgcgttggcttadcgagatgagtttccgcgcgatatacccgccaccttt	726



Db 601 TGTAAAGAACGCGGCTTGTTCAGC----- 625  
QY 727 cgactcatgctgagcgtctcccgctcgagcaattcgcccttcaatgggttaaacagca 786  
Db 626 -----CAGCAATTGCGCCCTTCAATGGGTTCAACAGCA 657  
QY 787 cgtctgaagtctcgaggcgcaaccgcatcacgttagatttgggcgagctctcagggcg 846  
Db 658 CGTCTGGAAGTTGGGGGCAACCCCGATCACGTTACGATTTGGGGGAGTCTCAGGGCG 717  
QY 847 aggttcggttatgaacagatattcgaaactgagocacccgaacgcatctccagcgga 906  
Db 718 AGGTCCGTTATGAACAGATCATTCGAAC----- 748  
QY 907 ctttcccccccccccccccccgctgacctccctctgtcttgaggggggaacacggtca 966  
Db 749 -----GGCGGCAACACCGTCA 764  
QY 967 aggtctcgtctcaagaagccctcttccacgctgpcatcgctcctcgtcttctctcc 1026  
Db 765 AGCTCTCGTCTCAAGAACCCCTCTTCCACGCTGCCATCGGCTCCTCGCTCTCTCC 824  
QY 1027 cctaccaagccaagtacaactcccccttcgagagctgtcttactccaaactcgtctcg 1086  
Db 825 CTTACCAAGCAAGTACAACTCCCTCTCGCGAGGTGCTACTTCCAACTGCTCTCGG 884  
QY 1087 cgacaaactcacaaagccgctcgtctcgtctgctcgaagctgtcgagcgtcg 1146  
Db 885 CGACAAACTGACAAAGCCGCTCGTCTGCTTTCGCTCGAAGCTGTGACGCTCGG 944  
QY 1147 cgtcgtcgtcgggcggtgaagaactcggcggttcccggttcgggttttggctgatg 1206  
Db 945 CGTCTCGTCTCGGGGCTGAAGACTCGCGCGGTTCCTCGGTTTGTGCTGATG 1004  
QY 1207 tcccggtcgtcgagggaccttcttgaactgagcgcggtcgtctcctctcgccaaggca 1266  
Db 1005 TCCGGTCTGTCGAGCGGACCTCTTACTAGCGCGCGGTGCTCTCTCGCAAGGCA 1064  
QY 1267 agaaacactcaatggtggtgagagctttcgaggttctcgatctcgtgacct 1326  
Db 1065 AGAAGAACCTCAATGGC----- 1081  
QY 1327 gtcgacggctcgagacacttctacggggtacaaacactcgacgaagatgattccc 1386  
Db 1082 -----AACCTCTACCGGATCAACAACTCGACGAAG----- 1115  
QY 1387 gtgagggctctgttgcgcagagagactgactgttcttttgcgaagattagattcat 1446  
Db 1116 -----GATTTCAT 1122  
QY 1447 attcactgagccactattcagaacacagatcagcagcagctcgagcgctctccca 1506  
Db 1123 ATTCACTGACCCACTATTGAGAACGACAGATCAGACACAGTCGCGCGGCTCTCCCA 1182  
QY 1507 gtctgacgcctctcgtccggtcttccctacatcacctcgagagcgccagggcgt 1566  
Db 1183 GTTCGACCGCTCTCTCGCGCTTTCCTCTATACCTCGAGGAGCGCCAGGCGT 1242  
QY 1567 cgcgaagcagtacccgattctcgagcgccggtcaaaagggcaaacaccttctctcgatctc 1626  
Db 1243 CGGGAAGCAGTACCGGATCTCCGACGCGCGGTCAAGGGCAACACCTCTCTCGCATCTC 1302  
QY 1627 ggcgtcatcgagactcgacctctgtgtggttccccgctgttcttcgagattccg 1686  
Db 1303 GGCGTATCGGACCTGACCTTCGT----- 1329  
QY 1687 ctgacttcccggttgcgcgagctgcccgcactactgacgcgagcggttcggtcggt 1746  
Db 1330 -----CTGCGCCGACCTACTTGACCGCGGAGCGGTTGCGCTCGT 1367  
QY 1747 ccgccccaaaggcctcttctgactaagcgcggtctaccacgagacgacactcgact 1806

Db 1368 CGCCCAAGAGGCGCTTTCGACTAGCGCGGCTCACCACGCGACCGACAACCTCGTACT 1427  
QY 1807 acatcgctccatctgaaacggcaaaagtcggtctcgtccgtccagctcttcgagcg 1866  
Db 1428 ACATCGGCTCCATCTGAGACGGCAAGATCGGTCTCGTCCGTCAGTCTTTCGAGCGG 1487  
QY 1867 cgtcggcggttcacatcgagacgttcaacccgcaacaaacgctgccaacaagaccatca 1926  
Db 1488 CGCTCGGCGCTTTCATCGAGACGTTCAACCCGAACAACAACGCTGCCAACAAGACCATCA 1547  
QY 1927 acccttactggcgagcttcgagctcggaagcagctcctcttcaacacacgagcgagg 1986  
Db 1548 ACCCTTACTGCGGACGTTTCGACTCGGCAAGCAGCTCTCTTCAACACGACGACGAGG 1607  
QY 1987 acacctctctcccgcgaccccgcatctgtgagacttcaagcttgaacgactttgca 2046  
Db 1608 ACACCTCTCTCCCGCGGACCCCGGATCGTTGAGACTTCAAGTTGACCGACTTTGGCA 1667  
QY 2047 cgagcagaagacaaagtcgagcttctggcggtggtcaatctcgtgaaacggtctct 2106  
Db 1668 CGAGCCAGAAAGACCAAGTGGGACTTCTGGGTGGGTCAATCTCGGTGAGCGGGTCTCT 1727  
QY 2107 agcggtctttc 2117  
Db 1728 AGCGCTCTTTC 1738

RESULT 5  
BD005938  
LOCUS BD005938 1738 bp DNA linear PAT 31-JAN-2002  
DEFINITION Cephalosporin esterase gene from Rhodospiridium toruloides.  
ACCESSION BD005938  
VERSION BD005938.1 GI:18634309  
KEYWORDS JP 2001501466-A/1.  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 1738)  
AUTHORS Polifino, M., Tonzi, S.M., Usher, J.J., Burnett, W.V. and Romanc, G.  
TITLE Cephalosporin esterase gene from Rhodospiridium toruloides  
JOURNAL Patent: JP 2001501466-A 1 06-FEB-2001;  
BRISTOL MYERS SQUIBB CO  
COMMENT OS Unidentified  
PN JP 2001501466-A/1  
PD 06-FEB-2001  
PE 11-SEP-1997 JP 1998514766  
PR 18-SEP-1996 US 60/026929  
PI MICHAEL POLIFINO, SEAN M TONZI, JOHN J USHER, WILLIAM V BURNETT,  
PI GUNA ROMANCIK  
PC C12P21/06, C12N1/00, C12N9/16, C12N15/00  
CC Strandedness: Both;  
Topology: Linear;  
FH Key Location/Qualifiers  
FT CDS Location/Qualifiers  
source 1..1738  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 327 a 639 c 444 g 328 t  
ORIGIN

Query Match 61.9%; Score 1375; DB 6; Length 1738;  
Best Local Similarity 84.7%; Pred. No. 3.6e-203;  
Matches 1738; Conservative 0; Mismatches 0; Indels 313; Gaps 5;

QY 67 actcgcgccatgctcttaacctttcaacctcttcaacctcgctcctcgtcgtcgagctccagct 126  
Db 1 ACTCGCGCATGCTCTTAACCTTCTACCTCTCTACCTCGCTCGCTCGCTCGCTCGCTCGCT 60  
QY 127 cgccttgcctctccgacctctctcgtccgacgacacacacacacacacacacacacacac 186  
Db 61 CGCCTTTCGCTCTCCGACCTCTCCCTCGTCCGCGCCGACCAACCAACGAGCCCTCCCT 120

187 cgtgaactcgtgctacgcccgcctaccaggctacttgaacgagagaccgcccgaactctactg 246  
121 CGTCGACCTCGCTACGCCGCTACCAAGGCTACTTGAACGAGACCGCGGACTCTACTG 180  
247 gtgacggaatcgtctacgctcgtcagctcagcttccaggctcctcagaagccgcgcgac 306  
181 GTGGCGGGGAATCCGCTACGGCTCGGCTACGCTTCCAGGCTCTCAGAGCGCCGGGAC 240  
307 gcacaagccgctcgcgaacgcagctgagtagtggaaccgatctgttggccggttagcgagg 366  
241 GCACAAGCCGTCGCAACGGGACTGAGTAGGACCGATCTGTGGCCGGTACGAGGG 300  
367 aaccaacacgaccgaagggcttgccgcgcctagcaacagctcgaagcgcgcgcgcagaa 426  
301 AACCAACACGACCAAGGGCTTGCCGCGCTTAGCAACAGCTCGAGAGCGCGCGCAGAA 360  
427 acagggcgtgagagattgctctctcctcaatgtcgttgcgccccgcgcgcgcgcgcg 486  
361 ACAGGCGTCGGAGATTGCTCTTCTCAATGTCTGTTGCCCGCGGCTCGTCGAGGG 420  
487 cgacaattctccgctcgtctacattcacggaggtggttacgccttcggcgatgcgag 546  
421 CGACAATCTCCGCTCTCTCTACATTACAGGAGGTGGCTACGCCCTTCGGCGATGGGAG 480  
547 caccggcagcgaacttgccgccttcacaaagcacacacggaacaaagatggttcgttga 606  
481 CACGGGAGCGACTTGCCGCGCTTACCAAGCACACAGGGAACCAAGATGTCGTTGATA 540  
607 tctcagtagcgtcgtcgcagcttctgctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 666  
541 TCTCAGTACGCTCGGACGCTTGTGTTCTCTGCTGGCCGCAAGCCATGAAGGACTACGG 600  
667 tgaacaacccgctgcttgaccaggtagtattcccgcatataccgcgcacattt 726  
601 TGTACAGAACCCCGCTGCTTGAC----- 625  
727 cgaactatgctgagcctctcccgcctcgcgagcaaatcgccttcaatgggttcaacagca 786  
626 -----CAGCAATTGCGCCCTTCAATGGGTTCAACAGCA 657  
787 cgtctgaagtctcggcgaaccccgatcaagttacagtttggggcgagctcgcagcgc 846  
658 CGTCTGAAGTCTCGGCGCAACCCCGATACGTTACGATTTGGGCGGAGCTCGAGCGC 717  
847 aggtctcgttatgaaccagatcattgcgaagtgagccacccgaacgcatctccagcga 906  
718 AGGTCTCGTTATGAACAGATCATTTGGAA----- 748  
907 cttcccccccccccccccccccgctgacctccctcgttcttgcagggcggaacaccgtca 966  
749 -----GGCGGCAACACCGTCA 764  
967 aggtctcgttctcaagaagccctcttccacgctgcacatggctcctccgcttctctcc 1026  
765 AGGTCTCGGTCTCAAGAAGCCCTCTTCCACGCTGCATCGGCTCTCCGCTCTCTCTCC 824  
1027 cctaccagcccaagtacaaactccccctcgcgcgagctgctctactccaaactcgtcgg 1086  
825 CCTACCAAGCCAGTACAACTCCCCCTCGCGGAGCTGCTCTACTCCCAACTCTCTCGG 884  
1087 gcacaaactgcacaaagccgctcgtcctcgttgcctcgaagctgtagcgtcgg 1146  
885 CGACAAATGACCAAGCCGCTCTGCTTCTGCTTGGCTTGGCTGCAAGCTGTCGACGCTCGG 944  
1147 cgtcgcgtcgcggggcggtgaagaactcggcggttcccgctcgttgggttttgggtcgtatg 1206  
945 CGCTCGTCTCGGCGGGCGGTGAAGAACTCGGCGGCTTCCCGCTTCCGCTTGGTTCGTATG 1004  
1207 tcccgctcgtcgcagcgcctcttgaactgagcgcgcgctcgtcctctctcgcgaaggca 1266  
1005 TCCGCGTCTGTCGACGGGACCTCTTGTGACTGAGCGCGGCTGCTCTCTCTCTCCCAAGGCA 1064

QY 1267 agaagaacctcaatggcgctgctggcgagagcttctcagtgcttcaggatctcgtgcacact 1326  
Db 1065 AGAAGAACCCTCAATGGC----- 1081  
QY 1327 gtcgaccggtcgcagaacactcttcacgggagatcaacaacactcgaagaatgagtcc 1386  
Db 1082 -----AACCTCTTACCGGGATCAACAACCTCGACGAAG----- 1115  
QY 1387 gtcgacggctctgttcgcccagcagactgactgttcttttggagaattacagatcat 1446  
Db 1116 -----GATTCAT 1122  
QY 1447 attcaactgaagccactattcaagaacagatcagagaccagtcgagcgcgtctccca 1506  
Db 1123 ATTCACCTGAGCCACTATTAGAAACGACACGATCAGGACCGACAGTCGAGCGCTCTCCCA 1182  
QY 1507 gttcgaacgctcctcgcggcctcttcccctacatcacctcgcgagagcgcagggcgt 1566  
Db 1183 GTTCGACCGGCTCTCGCGGCTTCTCCCTTACATCACCCTCGGAGGAGCCAGGCGT 1242  
QY 1567 cgcgaagcagtaaccgcatctcgcagcgcgcgtcaaaaggcaaaccttctctcgcattctc 1626  
Db 1243 CGCAAGCAGTACCGGATCTCCGACGCGCGTCAAAAGGCAACACTTCTCTCGCATCTC 1302  
QY 1627 ggcgctcagcgcgagcactcgtctgctgcttcccccgctgcttcttccgagattccg 1686  
Db 1303 GGCGGTATCGGAGCTCGACCTTCGT----- 1329  
QY 1687 ctgacttccgcttgcgcgcagctgcccgaactactgagacgcgagcgcttcgctcgt 1746  
Db 1330 -----CTGCCGACTTCTGACCGCGAGCGCTTCGGCTCGT 1367  
QY 1747 ccgcccacaagcgctcttcgactacgcgcgcgtccaccgcgacgcgacactcglact 1806  
Db 1368 CCGCCCAACAGGGCTCTTTCGACTACGGCGCGGCTCACCAGCCGACCGACAACTTCGTACT 1427  
QY 1807 acatcgctcctcatctggaacggcgaagtgctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 1866  
Db 1428 ACATCGGTCTCATCTGGAACGGCAAGTCTGGTCTCGTCTCGTCTCAGTCTCTGACGCG 1487  
QY 1867 cgtcgcgggcttcatcgagacgttcaaccgcgaacaacaacagcgtgcgaacacatca 1926  
Db 1488 CGTCTCGCGGCTTCATCGAGACGTTCAACCCGAACAACAGCTGCGCAACAGACCACTCA 1547  
QY 1927 accttactgcccagcgttcgactcgggcaagcagctcctcttcaaacacgacgacgagg 1986  
Db 1548 ACCCTTACTGGCCGACGTTTCGACTCGGCAAGCAGCTCTCTTCAACACGACGACGAGGG 1607  
QY 1987 acacctctctcgcgcgacccgcgcatcgttgagacttcaagcttaccgactttggca 2046  
Db 1608 ACACCTCTCTCCCGCGACCCCGGCATCGTTGAGACTTCAAGCTTCAAGCTTACCAGCTTTGGCA 1667  
QY 2047 cgaacccaagaacaaagtgcgacttctggcggtggtggtggtggtggtggtggtggtggt 2106  
Db 1668 CGAGCCAGAGAACAAAGTGGGACTTCTGCGCTGGGTCAATCTCGGTGAACCGGGTCTCT 1727  
QY 2107 aggcgtctttc 2117  
Db 1728 AGGCGTCTTTC 1738

## RESULT 6

AF025410 LOCUS 1738 bp mRNA linear PLN 30-DEC-1997  
DEFINITION Rhodospiridium toruloides cephalosporin esterase precursor, mRNA,  
complete cds.  
ACCESSION AF025410  
VERSION AF025410.1 GI:2731567  
SOURCE Rhodospiridium toruloides.  
ORGANISM Rhodospiridium toruloides  
Eukaryote; Fungi; Basidiomycota; Urediniomycetes;  
Microbotryomycetidae; Heterogastriales; Sporidiolaceae;





```
QY 1277 caatggcgtgctggcgagcttttcagagtgcttcaggatctcgtgacactgtcgacgcgc 1336
|||||
Db 1065 CAATGGC----- 1071
QY 1337 tcgcagaacctttcaccggaggaatacaacaacctgcagaaagatagttcccgctgcagcgct 1396
|||||
Db 1072 -----AACCTCTTCACCGGGATCAACAACTCGACGAAG----- 1105
QY 1397 ctgttgcgccagcgagactgactgttcttttgcgaagatcagcattcatattcactgac 1456
|||||
Db 1106 -----GATCATATTCACTGAC 1122
QY 1457 gccactattcagaacacacagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1516
|||||
Db 1123 GCCACTATTCAAGACGACAGATCAGGACAGTCGCGAGCGCTCCACAGTTCCGACCGC 1182
QY 1517 ctctgcgcgcgtcttccctccatcatacctcagagagcagcagcagcagcagcagcagcagc 1576
|||||
Db 1183 CTCTCGCGCGCTCTTCCCTTACATCACCCTCGAGAGCGCGCAGCGCGTCCGGAAGCAG 1242
QY 1577 taccgatctccagcgcgcgtcaagggaacacaccttctcgcacatcgcgcgtcctc 1636
|||||
Db 1243 TACCGATCTCCGACGCGCGCTCAAGGGCAACACCTTCTCTCGCATCTCGCGCGTCTC 1302
QY 1637 gcggaactcgacctgtgtgtgttcccccgtctcttccagagtattccgctgactccc 1696
|||||
Db 1303 GCGGACTCGACCTCGT----- 1319
QY 1697 gcttgccgcagctgcccgaactactgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1756
|||||
Db 1320 -----CTGCCGACCTACTGGACCGCGGAGCGTTCGGCTCGTCCGCGCCACAA 1367
QY 1757 gggcctctcgaactaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1816
|||||
Db 1368 GGGCTCTTGCATACGCGCGCTCACACGCGGACGACCACTCGTACTACATCGGCTC 1427
QY 1817 catctgaacgcgaagtcggtctcgtctcgtctcgtctcgtctcgtctcgtctcgtctcgt 1876
|||||
Db 1428 CATCTGAACGGCAAGATCGGTCTCGTCTCGTCTCGTCTCGTCTCGTCTCGTCTCGTCTCG 1487
QY 1877 ctteatgcagcgttcaaccggaacaaacagctgcgaacagacacacacacacacacacac 1936
|||||
Db 1488 CTTTCATCGACGCTTCAACCGCAACAAACGCTGCGCAACAGACCACTCAACCTTACTG 1547
QY 1937 gccgacttcgactcggcgaagcagctcctcttcaacacagcagcagcagcagcagcagc 1996
|||||
Db 1548 GCGGACGTTGACTCGGCGAAGCAGCTCTCTTCAACACGACGACGAGGACACCCCTCTC 1607
QY 1997 tccgcgcgacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2056
|||||
Db 1608 TCCGCGGACCGCGGCATCGTTGAGACTTCAAGCTTGACCGACTTGGCAGGAGCCAGAA 1667
QY 2057 gaccaatgcgactctcggcgtgggtcaactcgcgtgaacgcgggtctc 2105
|||||
Db 1668 GACCAAGTGCAGCTCTGCGGTGGGTCAATCTCGTGAACGCGGGTCTC 1716

RESULT 8
AC087563/c
LOCUS 298166 bp DNA linear HTG 10-JAN-2001
DEFINITION Homo sapiens chromosome 16 clone RP11-358F6, WORKING DRAFT
SEQUENCE, 75 unordered pieces.
ACCESSION AC087563
VERSION AC087563.1 GI:12061429
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 298166)
REFERENCE
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
JOURNAL Unpublished
```

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

2 (bases 1 to 298166)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (10-JAN-2001) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov  
-----

Project Information  
Center Project Name: 0  
Center clone name: RPCI-11\_358F6  
-----

Summary Statistics  
Consensus quality: 189570 bases at least Q40  
Consensus quality: 207567 bases at least Q30  
Consensus quality: 217222 bases at least Q20  
Estimated insert size: 231000; agarose-fp estimation  
Quality coverage: 8.11 in Q20 bases; agarose-fp estimation  
Quality coverage: 6.44 in Q20 bases; sum-of-contigs estimation.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 75 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1136: contig of 1136 bp in length  
1137: gap of unknown length  
1237: contig of 1021 bp in length  
2258: gap of unknown length  
2357: gap of unknown length  
3495: contig of 1138 bp in length  
3595: gap of unknown length  
4702: contig of 1107 bp in length  
4802: gap of unknown length  
5944: contig of 1142 bp in length  
6044: gap of unknown length  
7153: contig of 1109 bp in length  
7253: gap of unknown length  
8400: contig of 1147 bp in length  
8500: gap of unknown length  
9505: contig of 1005 bp in length  
9605: gap of unknown length  
10744: contig of 1139 bp in length  
10844: gap of unknown length  
12291: contig of 1447 bp in length  
12391: gap of unknown length  
13519: contig of 1128 bp in length  
13619: gap of unknown length  
15101: contig of 1482 bp in length  
15201: gap of unknown length  
16501: contig of 1300 bp in length  
16601: gap of unknown length  
17767: contig of 1166 bp in length  
17867: gap of unknown length  
19234: contig of 1367 bp in length  
19334: gap of unknown length  
20427: contig of 1093 bp in length  
20527: gap of unknown length  
21623: contig of 1096 bp in length  
21723: gap of unknown length  
23056: contig of 1333 bp in length  
23156: gap of unknown length  
24305: contig of 1149 bp in length  
24405: gap of unknown length  
25647: contig of 1242 bp in length  
25747: gap of unknown length  
27088: contig of 1321 bp in length  
27168: gap of unknown length  
28319: contig of 1151 bp in length



```
QY 459 tcgtgccccgcgcgtctgcgaggcgacaattctccgtctctctgattacattcaacg 518
      |||||
Db 18145 NNNNNCCCCCCCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 18086
QY 519 gagtggtacgctctcggcgatgcgagcaacggcgagcacttgcgcgttcaacaagc 578
      |||||
Db 18085 NNNNNNNCCCCCCCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 18026
QY 579 acacgggaacaagatggtgtgtgttaattccagtaaccgtctcgcgagcttggttcc 638
      |||||
Db 18025 NNNNNCCCCNNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 17966
QY 639 tcgtggtcgaagcatgaagtaagtgtaacgaacgcgcggtgtcttgaccagtgga 698
      |||||
Db 17965 NAGTGCCTATTCAAGGNNNTTTCANGACNNCTNNCAACGCGCNCCTTACAAGCTGTTA 17906
QY 699 gttcccgatgataccgcgcacacttccagtaactgactcatgtaacgcctctccgcgcagc 758
      |||||
Db 17905 AAANNCCNCCGCGCCNCCGCGCCNCCGCGCCNCCNCCNCCNCCNCCNCCNCCN 17846
QY 759 aattcgccctcaatgggttcaacagcacgtctcgaagtctcgcgcaaccccgatcacg 818
      |||||
Db 17845 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 17786
QY 819 ttacgattggggcgagctgcaggcgaggttcgttatgaaccagatcattgcgaacg 878
      |||||
Db 17785 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 17726
QY 879 tgagccaccgaaccgatctccagcgaattctcccccgcgcgcgcgcgcgcgcgcgc 938
      |||||
Db 17725 CCGCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 17666
QY 939 ctgctcttgaggcggaacacgcgtcaaggctcgtctcgaagagccctcttccac 998
      |||||
Db 17665 CCCCCCCCCNNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 17606
QY 999 gctgcatacgtctcgtctctctctctctctctctctctctctctctctctctctc 1058
      |||||
Db 17605 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 17546
QY 1059 gactgtctactcccaactcgtctcgcgcgacaaactgcacaaagcgcctctcttc 1118
      |||||
Db 17545 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 17486
QY 1119 gcttgctcgaagctgcagctgcgcgcgtcgtcgcgcgcgtggaagaactcgcgcg 1178
      |||||
Db 17485 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 17426
QY 1179 gcgtcccgctcgggttttggtgtatgtcccggtgctgcaggggacatttctgactgag 1238
      |||||
Db 17425 CCCCCCCCCNNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 17366
QY 1239 cgcgcgtcgtctctctcgaagggcaagaaacactcaatggcgtgctggtgcgagctt 1298
      |||||
Db 17365 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 17306
QY 1299 tcgagtgcttcaggtatcgtgcactgtgcagcggctgcagaactctt-----cac 1353
      |||||
Db 17305 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 17246
QY 1354 cgggatacaaacctcgaacgaagatgagttcccgtagcggctgtctggtcgcagcagaga 1413
      |||||
Db 17245 CCCCCCCCCNNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 17186
QY 1414 ctgactgtcttttggaagattacgattcatattcactgacgcacattatcagaacga 1473
      |||||
Db 17185 NNNNNCCCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 17126
QY 1474 cactgacgacacagtcgcagcgcgtctccagttcagtcagcgcctctcgcgcgcctctt 1533
      |||||
Db 17125 CNCNNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 17066
QY 1534 cccctacatacctcgcgagggcgccagcgcgtcgcggaagcagtcacgcgtctcgcgcgcg 1593
```

## RESULT 9

```
AC087190/c 165556 bp DNA linear HTG 25-APR-2001
LOCUS Homo sapiens chromosome 16 clone RP11-47311, WORKING DRAFT
DEFINITION
AC087190
ACCESSION AC087190
VERSION AC087190.2 GI:13786429
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 165556)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 165556)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (13-DEC-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Apr 25, 2001 this sequence version replaced gi:11693350.
COMMENT
-----Genome Center
Center: Joint Genome Institute
Center code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 0
Center clone name: RPCI-11-47311
-----
Summary Statistics
Consensus quality: 112073 bases at least Q40
Consensus quality: 126903 bases at least Q30
Consensus quality: 135461 bases at least Q20
Estimated insert size: 138300; agarose-fp estimation
Quality coverage: 5.37 in Q20 bases; agarose-fp estimation
Quality coverage: 4.56 in Q20 bases; sum-of-contigs estimation
* NOTE: This is a 'working draft' sequence. It currently
* consists of 28 contigs. The true order of the pieces
```









```
* * 60548: gap of unknown length
* * 60648: Contig of 1096 bp in length
* * 61743: gap of unknown length
* * 61843: gap of unknown length
* * 62998: Contig of 1155 bp in length
* * 63098: gap of unknown length
* * 64220: Contig of 1122 bp in length
* * 64320: gap of unknown length
* * 65398: Contig of 1078 bp in length
* * 65498: gap of unknown length
* * 66762: Contig of 1264 bp in length
* * 66763: gap of unknown length
* * 66863: Contig of 1163 bp in length
* * 68026: gap of unknown length
* * 68126: Contig of 1275 bp in length
* * 69401: gap of unknown length
* * 69501: Contig of 1218 bp in length
* * 70718: gap of unknown length
* * 70818: Contig of 1177 bp in length
* * 71995: gap of unknown length
* * 72096: Contig of 1137 bp in length
* * 73233: gap of unknown length
* * 73333: Contig of 1144 bp in length
* * 74477: gap of unknown length
* * 74577: Contig of 1287 bp in length
* * 75864: gap of unknown length
* * 75964: Contig of 1058 bp in length
* * 77022: gap of unknown length
* * 77122: Contig of 1091 bp in length
* * 78213: gap of unknown length
* * 78313: Contig of 1112 bp in length
* * 79424: gap of unknown length
* * 79524: Contig of 1093 bp in length
* * 80618: gap of unknown length
* * 80718: Contig of 1087 bp in length
* * 81804: gap of unknown length
* * 81904: Contig of 1103 bp in length
* * 83008: gap of unknown length
* * 83108: Contig of 1060 bp in length
* * 84168: gap of unknown length
* * 84268: Contig of 1080 bp in length
* * 85347: gap of unknown length
* * 85447: Contig of 1074 bp in length
* * 86521: gap of unknown length
* * 86622: Contig of 1123 bp in length
* * 87745: gap of unknown length
* * 87845: Contig of 1121 bp in length
* * 88966: gap of unknown length
* * 89066: Contig of 1111 bp in length
* * 90177: gap of unknown length
* * 90277: Contig of 1003 bp in length
* * 91280: gap of unknown length
* * 91380: Contig of 1184 bp in length
* * 92584: gap of unknown length
* * 92684: Contig of 1025 bp in length
* * 93689: gap of unknown length
* * 93789: Contig of 1146 bp in length
* * 94935: gap of unknown length
* * 95035: Contig of 1077 bp in length
* * 96112: gap of unknown length
* * 96212: Contig of 2141 bp in length
* * 98353: gap of unknown length
* * 98453: Contig of 1193 bp in length
* * 99646: gap of unknown length

Query Match 3.38; Score 73.4; DB 2; Length 303091;
Best Local Similarity 21.58; Pred. No. 0.013;
Matches 239; Conservative 0; Mismatches 872; Indels 1; Gaps 1;

Qy 5 ccaccgaactgtccggtttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcc 64
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 58104 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 58045

Qy 65 cgactgcgcgcatgtcttcttcttcttcttcttcttcttcttcttcttcttcttcttcc 124
```

```
Db 58044 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 57985
Qy 125 ctgcctttgctctcgcgacctcctcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 184
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 57984 CCCCCCNCNNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 57925
Qy 185 gtctgcgacctcgcgtacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 244
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 57924 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 57865
Qy 245 tggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 302
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 57864 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 57805
Qy 303 cgacgcacaaaggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 362
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 57804 ANCCCCCCCCGNNCCTCCCCCAANNNNNNCCAGNNNNNTCCATGNGTCATTCCGGNAGNNCTCG 57745
Qy 363 aggaacacacacacacacacacacacacacacacacacacacacacacacacacacac 422
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 57744 TTCGNACCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 57685
Qy 423 agaaacaggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 482
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 57684 TNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 57625
Qy 483 agggcgacaatctccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 542
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 57624 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 57565
Qy 543 cgagcaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 602
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 57564 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 57505
Qy 603 taaatctcagtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 662
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 57504 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 57445
Qy 663 acggtgtaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 722
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 57444 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 57385
Qy 723 ctctgcactcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 782
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 57384 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 57325
Qy 783 agcagctcgaagtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 842
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 57324 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 57265
Qy 843 gcgcagggtcgttatgaacacagatcattgcgaagtcgcgcgcgcgcgcgcgcgcgcgc 902
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 57264 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 57205
Qy 903 cgcactttccccccccccccccccccccccccccccccccccccccccccccccccccc 962
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 57204 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 57145
Qy 963 gtcaaggctcgtgtctcaagaagccctcttccacgtgcgcgcgcgcgcgcgcgcgcgc 1022
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 57144 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 57085
Qy 1023 ctccctaccaagcaagtaacaactcccccttcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1082
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 57084 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 57025
Qy 1083 tcggcgacaaactgcacaaagccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1115
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 57024 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 56992
```

RESULT 11



Qy 65 ggaatcgccgcatgctcttaaacctcttcaacctctgctcctcctcctgctgagcagctccag 124  
Db 7244 CC 7185  
Qy 125 ctcgcttctgctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 184  
Db 7184 CC 7125  
Qy 185 gtgctgacctcggctacgctcctcctcctcctcctcctcctcctcctcctcctcctcctc 244  
Db 7124 CC 7065  
Qy 245 tgggtggcggaatcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 304  
Db 7064 CC 7005  
Qy 305 acgcaaacgctcgcgcaagcgaactgagtgatggacgatctgttgccgctgagcag 364  
Db 7004 NNN 6945  
Qy 365 ggaacaaacagacaaagggcttgcgcgcgcctagacagctcagcagcgcgcgcag 424  
Db 6944 NNNAAACANNCCNNNNNNNNCCCGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6885  
Qy 425 aaacaggcgtcggaggattgctctctcctcctcctcctcctcctcctcctcctcctc 484  
Db 6884 CGGTCTCTGTNNANANNNGNNTCTCCCATGGGGTCTCGATGTTNNTACNACN 6825  
Qy 485 ggcgaacatctccgctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 544  
Db 6824 CNGCCCCCGNCTNNNNNNNNCCCTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6765  
Qy 545 agcaccgagcagcttgcgcctcctcctcctcctcctcctcctcctcctcctcctcctc 604  
Db 6764 NNN 6705  
Qy 605 aatctccagtcacgtcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 664  
Db 6704 NNN 6645  
Qy 665 ggttaacgaacgcttgcctcctcctcctcctcctcctcctcctcctcctcctcctcct 724  
Db 6644 NNN 6585  
Qy 725 ttcgactcatgctgacgctcctcctcctcctcctcctcctcctcctcctcctcctcct 784  
Db 6584 NNN 6525  
Qy 785 cactctcgaagtctcggcgcaaccccgatcacgttactgagtgagtgagtgagtgagtc 844  
Db 6524 NNN 6465  
Qy 845 gcagggtcgttatgaacagatcattgcgaactgagcaccgcgaacagcagcagcagc 904  
Db 6464 NNN 6405  
Qy 905 gacttctcccccctcccccctcccccctcctcctcctcctcctcctcctcctcctcct 964  
Db 6404 CC 6345  
Qy 965 caaggtctcgtctcgaagccctctcctcctcctcctcctcctcctcctcctcctcctcct 1024  
Db 6344 CC 6285  
Qy 1025 cccctacaaacgaagtacaaactccctctcctcctcctcctcctcctcctcctcctcct 1084  
Db 6284 CC 6225  
Qy 1085 ggcgacaaactgcacaaagcgcctcctcctcctcctcctcctcctcctcctcctcctc 1127  
Db 6224 CC 6182

RESULT 12  
AC090552/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

AC090552 181988 bp DNA linear HTG 01-AUG-2001  
Homo sapiens chromosome 3 clone RP11-759M11, WORKING DRAFT  
SEQUENCE, 19 unordered pieces.  
AC090552  
AC090552.2 GI:13786433  
HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 181988)  
DOE Joint Genome Institute.  
Sequencing of Human Chromosome 3  
Unpublished  
2 (bases 1 to 181988)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (03-MAR-2001) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
On Apr 25, 2001 this sequence version replaced gi:13194218.  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: <http://www.jgi.doe.gov>  
-----

Project Information  
Center Project Name: 0  
Center clone name: RPC1-11\_759M11  
-----

## Summary Statistics

Consensus quality: 140511 bases at least Q40  
Consensus quality: 156778 bases at least Q30  
Consensus quality: 164746 bases at least Q20  
Estimated insert size: 140300; agarose-fp estimation  
Estimated insert size: 180188; sum-of-contigs estimation  
Quality coverage: 4.92 in Q20 bases; agarose-fp estimation  
Quality coverage: 3.83 in Q20 bases; sum-of-contigs estimation.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 19 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1372: contig of 1372 bp in length  
\* 1373 1472: gap of unknown length  
\* 1473 2486: contig of 1014 bp in length  
\* 2487 2586: gap of unknown length  
\* 2587 4020: contig of 1434 bp in length  
\* 4021 4120: gap of unknown length  
\* 4121 5414: contig of 1294 bp in length  
\* 5415 5514: gap of unknown length  
\* 5515 6649: contig of 1135 bp in length  
\* 6650 6750: gap of unknown length  
\* 6750 7896: contig of 1147 bp in length  
\* 7897 9226: contig of 1230 bp in length  
\* 9227 9327: gap of unknown length  
\* 9328 10600: contig of 1274 bp in length  
\* 10601 10700: gap of unknown length  
\* 10701 12783: contig of 2083 bp in length  
\* 12784 12884: gap of unknown length  
\* 12884 14419: contig of 1536 bp in length  
\* 14420 14519: gap of unknown length  
\* 14520 18200: contig of 3681 bp in length  
\* 18201 18300: gap of unknown length  
\* 18301 23250: contig of 4950 bp in length  
\* 23251 23351: gap of unknown length  
\* 23351 28611: contig of 5261 bp in length

```
* 28612 28711: gap of unknown length
* 28712 35514: contig of 6803 bp in length
* 35515 35614: gap of unknown length
* 35615 49317: contig of 13703 bp in length
* 49318 49417: gap of unknown length
* 49418 69186: contig of 19769 bp in length
* 69187 69286: gap of unknown length
* 69287 98146: contig of 28860 bp in length
* 98147 98246: gap of unknown length
* 98247 127092: contig of 28846 bp in length
* 127093 127193: gap of unknown length
* 127193 181988: contig of 54796 bp in length.
FEATURES
  source
    1. 181988
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="3"
      /clone="RP11-759W11"
      /clone_lib="RPC1 human BAC library 11"
BASE COUNT 53222 a 36221 c 36728 g 51237 t 4580 others
ORIGIN
Query Match 3.18; Score 69.8; DB 2; Length 181988;
Best Local Similarity 20.5%; Pred. No. 0.051;
Matches 231; Conservative 0; Mismatches 894; Indels 3; Gaps 1;
QY 5 ccaccgaactctccgcttttggctttcttcttcttcttcttcttcttcttcttcttcttcc 64
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3185 CCNNNCCGCCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 3126
QY 65 cgactgcgcgactctcttaactcttcttcttcttcttcttcttcttcttcttcttcc 124
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3125 NCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCC 3066
QY 125 ctgcgttgcctcttcgactcttcttcttcttcttcttcttcttcttcttcttcttcc 184
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3065 CCCCNCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCC 3005
QY 185 gtcgtgactctgcctacgcgcgtacaaagctactgactgactgactgactgactctac 244
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3005 CNNCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCC 2946
QY 245 tggtagcggaatcgcgtacgctcgcgtcgcgtcgcgtcgcgtcgcgtcgcgtcgcgcg 304
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2945 CCCCNCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2886
QY 305 acgcacaaggcgtccgcacagcactgactgactgactgactgactgactgactgactgag 364
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2885 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2826
QY 365 ggaaccaacagcacaagggttcgcgcgcgtcagcaacagcgtcagcagcgcgcgcgcag 424
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2825 NCCGCCNCCGCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2766
QY 425 aaacaggcgtcggaggattgccttctctcaatctcgttgcgtccgcgcgcgcgtcgtgcg 484
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2765 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2706
QY 485 ggcgacaattcttcgctctctctctctctctctctctctctctctctctctctctctcgcg 544
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2705 NCCGCCNCCGCCNAAANNNNNNNGGGGTTCGTGNGGGGCGTGTGTGTTCTGTCGGCGCN 2646
QY 545 agcaccgacgactttgcgccttccacagcacacagcagcagcagcagcagcagcagcagc 604
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2645 CGNNMCGCTNNNNNGCTCTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2586
QY 605 aatctccagtagctctcgcgagctttgtttctctcgtcgtgcagcagcagcagcagcagc 664
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2585 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2526
QY 665 ggtgtaacgaacgcgcgttcgttgaccaggtgagtttccgcgactgataccgcgcacact 724
```

```
Db 2525 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2466
QY 725 ttgcactcatgctgaagcctctccgctcgcagcaattcgccttcaatgggttcaacag 784
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2465 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2406
QY 785 cactgtctgaagtgcggcgcaaccccgatcacgttaagatttggggcgagctcgcagcc 844
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2405 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2346
QY 845 gagggtcgttatgaaccagatcatgcaacgtgagccaccgacccgacgactccagcc 904
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2345 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2286
QY 905 gactttcccccccccccccccccccccccccccccccccccccccccccccccccc 964
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2285 NNNNNNNCCGCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 2229
QY 965 caaggtctcgtctcaagaagccctcttccacgctgcgcgtcgcgtcgcgtcgcgtctct 1024
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2228 CCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCC 2169
QY 1025 cccctacaagcgaagtacaaactcccccttcgcgcgagctgcttactctcccaactcgtctc 1084
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2168 CCGNNNNNNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 2109
QY 1085 ggcgacaaactgcacaaagcgcgtcgtcgtcgtcgtcgtcgtcgtcgtcgaagc 1132
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2108 CNGCCGCCNCCGCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 2061
RESULT 13
LOCUS AP001552 150120 bp DNA linear PLN 16-JUN-2000
DEFINITION Oryza sativa genomic DNA, chromosome 6, PAC clone:P0029D06.
ACCESSION AP001552
VERSION AP001552.1 GI:7363267
KEYWORDS
SOURCE
  ORGANISM
    Oryza sativa (cultivar:Nipponbare) DNA, clone:P0029D06.
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
    Ehrhartoideae; Oryzaceae; Oryza.
  AUTHORS
    Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE
    Oryza sativa nipponbare(GA3) genomic DNA, chromosome 6, PAC
    clone:P0029D06
  JOURNAL
    Published Only in Database (2000) In press
  REFERENCE
    2 (bases 1 to 150120)
    Sasaki,T., Matsumoto,T. and Yamamoto,K.
  AUTHORS
    Direct Submission
  TITLE
    Submitted (29-MAR-2000) Takuji Sasaki, National Institute of
    Agrobiological Resources, Rice Genome Research Program; Kannondai
    2-1-2, Tsukuba, Ibaraki 305-8602, Japan
    (E-mail:tsasaki@abrr.affrc.go.jp,
    URL:http://www.dna.affrc.go.jp:82/, Tel:81-298-38-7441,
    Fax:81-298-38-7468)
  COMMENT
    The orientation of the sequence is from SP6 to T7 of the PAC clone.
    Genes were predicted from the integrated results of the
    following:GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as
    SplicePredictor (October 1998 version). The genomic sequence was
    searched against the non-redundant database NRP (PIR, SWISSPROT,
    GENPEPT, PDB) from MAF DNA bank and the cDNA sequence database at
    RGP. Protein similarities of the coding regions were searched
    against NRP with BLASTP2.0. ESTs represent the identified cDNA
    sequences using BLASTN2.0 with the corresponding DDBJ accession no.
    and RGP clone ID.
    This sequence of this clone has an overlap with P0541H01 clone,
    DDBJ:AP001389 at the 3' end.
    This clone ends at the position 26,826 of P0541H01. Detailed
    information on overlap and assembly quality together with
    annotation of this entry at
    http://www.dna.affrc.go.jp:82/genomicdata/GenomeFinished.html.
```

FEATURES	Location/Qualifiers	source
CDS	<p>1. 150120  /organism="Oryza sativa"  /cultivar="Nipponbare"  /db_xref="taxon:4530"  /chromosome="6"  /clone="P0029D08"  complement(join(1887..2096,2316..2390))  /note="hypothetical protein"  /codon_start=1  /protein_id="BAA93012.1"  /db_xref="GI:7363268"  translation="MIKAPITAVTLGRSRLCEGLAAGRGDCANRVSGLCPADVE  RKVEVAAGIEGSRCLGAEGLVSCSWKAGLAWPAAGGGLARHEA"  join(2761..3848,5363..5544,5878..5959,6846..7125)  /note="hypothetical protein"  /codon_start=1  /protein_id="BAA93013.1"  /db_xref="GI:7363269"  translation="MARRKGKOLISOLPDDILLHILSMVRYKEAVTAASVREWKHLH  TKLPALFSFVULGAQSSLSOSRORVDSMARTLRRACGPDRTVYKRLCLAYKDY  PMCHRGADFIALAAASSLGLFNCNPKNLRNDDAGPWSLHLPAATACLSSMESCWYSVR  PVRHVGFGSALKSLTFKDSFWLHPGYLQDTAPPSLEELHISGCTLSGSIETISATM  PRKHLRIADVSVSLGTAIAIADLTLVPSCHDGKGPDPSPSHMLCVETLFR  ASPTYESYELRAPKLRFVDWRCYAKEVRVAVGRLSDVVLPAFLPCYNEAK  REFLOMEOCKLMDILOGIMPGRWKYVQSLPYPRANKVRWGQDDREBEVEVQWHPR  MDMLTTMAASAGPGGQCTAASGCVYMAATARAGEDGNSRLGSGYMNEDUGIGCAARW  RADRARREMRPGGQRRGGRNEAVTGWARRRAERYSIVTAGGSTSARLRRPDAAR  TSGSAHEPDDRGAGAEGBEGAGSFSYKFKON"  join(8143..8539,8557..8908,9092..9284)  /note="hypothetical protein"  /codon_start=1  /protein_id="BAA93014.1"  /db_xref="GI:7363270"  translation="MVVDKRPAPVPLARRPASRCRLDRRYAEQLPPPRKGGAPAG  EGPVSREASSAGAGASHICPCSSPTLLCKSGTPPSGVGRERESGIGEMKSM  APTSVELLSAGSAGEVELAGSGPSLLTGKSEVAVAHGARGHRIQAAVYSMA  TPGPDGHRVPPADYTRGIRREAPASGPDVVCAMRRRGLVAVAAVPPPPRP  TPRRHLPLTPNHRHRYPCWLRSRREGKRAAAEVEEQRAAETPLPASEVRENK  SDHRAHGLRRRGQGRVPRAGHRTMAVTVGEH"  join(11170..11233,13292..13386)  /note="hypothetical protein"  /codon_start=1  /protein_id="BAA93015.1"  /db_xref="GI:7363271"  translation="MNLQVHPKRGVANGSDDKGILDGSDETNLFETSTKDFNS  GAGKREQD"  join(13886..13953,14703..14914,14920..15114,15144..15314,  15374..15492)  /note="hypothetical protein"  /codon_start=1  /protein_id="BAA93016.1"  /db_xref="GI:7363272"  translation="MNEADQPSPADPTSPRPLPGADYKSGGKDMRAPLYSGSRE  RGDAGTQWPLAVAGRCGGWAGTGHGSKREGASAGLSAPSPAGERRRQAT  DGDGDRDTSRRGRHRIATRMGAARAREGEGVLTVGPRATREITTINGDAA  ELQDTRLEEMREVRGREGVHGEREPMTAAAMEGLGTRETTFADPGGIRWRGRRAAA  SGGRSGKGGGNDRGHGRWRKEEGKEGCEAL"  join(19096..19172,20123..20329,20389..20395)  /note="hypothetical protein"  /codon_start=1  /protein_id="BAA93017.1"  /db_xref="GI:7363273"  translation="MQIETPHQCLVQWQDKIEIIPANRPQCSRRRLGTAAPTRRP  DIPLPSPSRRTPLAASAAAPSRPPRASPGRHSIGSGSALRLSGLRI"  21587..21769  /note="5' LTR"  join(21843..22489,22572..23072,23168..24293)  /note="Similar to Arabidopsis thaliana DNA chromosome 4,  BAC clone F17a8; putative protein. (AL049482)"  /codon_start=1  /protein_id="BAA93018.1"  /db_xref="GI:7363274"</p>	LTR
CDS	<p>complement(join(29217..30334,31321..31609,31701..31862))  /note="hypothetical protein"  /codon_start=1  /protein_id="BAA93019.1"  /db_xref="GI:7363275"  translation="MIQIKIKKKRDEVLVATRLGKVSLSQALRMDGRVSGDDADGE  RPLRRRAQCQPKSKVGLDWLGLPMIITSLTHTPPRLASTPIGRAAAAAVATAGL  VSPAPHGKLRVSKRSGTGAAGSGEDGHRRRRSPQRSPSLSAKGLMSMLHGI  IALSSFDQLAFSGTCSWRAALSFPSTIYTFPPHLKPDIPNSHLCSSFRVTL  LYKCRWQDGPDKRTLSLCSAPQNTPNMRVILGCSYGVLYIESYENCLLVDMYTGAK  VYPKLQSGAKNETIYILTAPLNFISHLILCSRSIYFWQVGVSHSEHPFGGRI  LQVTLQGEFFAMDFHRLTRMFAPQLSQEYGVVVGEMFVGVHFKPWLVTISGDM  LMDLSVGLHSYFPGTFOVFLDFSAQTAKMMKELNSALFSLDRRNFTFCT  NPERWGGKNCIYVAKPSDESDPTAVELQGPVPGATHCVYPYSHPLLRTEGCSOLE  YLWVLPSTFNGVDQ"  32671..33849  /note="hypothetical protein"  /codon_start=1  /protein_id="BAA93020.1"  /db_xref="GI:7363276"  translation="MOSPSKSEASGAPATSVIEGMAELPEGLLHSIVALLGSLFD  LAFPTGCHSWRAAFSSYPSKSNFRILPPLVPRVVRKAPYSNNGHRLRSCVID  LNANRPLKCIQPEITQRHMFAGSHGQICCRGYCLVVDVFTGAEVSPPLPSE  NCDPEYGLITAPITSTOSHLIISTOSLDFWPGSDSWSLKLPRVNDVIFENG  OLIAVIEYSLYTLQAPILRLKIKTLMMDNNECPYMPFVYVCDMLLIVDHYISF  SFGAPVYRPLRDMTKPAKVEKKNLWALFIGDARSPFSCFNPRMGRGNC  LYAHYSQPSLSLHGLDADAVDNDNLVFKRWYRQLQALWVYPSMFYSAGDQ  complement(join(34717..34935,35401..35912,36242..36363,  36470..36542,36626..36754,36936..37247,37336..38617,  38724..39212))  /note="Similar to Arabidopsis thaliana DNA chromosome V,  BAC clone TM021B04; N. tabacum membrane-associated  salt-inducible protein. (AF007271)"  /codon_start=1  /protein_id="BAA93021.1"  /db_xref="GI:7363277"  translation="MAAIIAGAAATAVSTICSYSYEDDGAATSWLSGTSSTSSSS  SORORYRLLHDEACRLRRERGOGSAGTPRWRTTDOMARYVEDDRAGHVGRH  VVAVAAARATASCSPSSADMRANASFAKTIFREKCEVLHELHGWGRFPAW  KLQCEPVSVAITILLRLQGVQKVLAEVTFLEMLQACEPDVACGILLACAYAK  KGLNDMLFYAARRDIVPSISVFNFMSSQKQLHGVILHQBOMLEANYAPNOF  TYTVVIGSAKEGMELEADAFEMRRRREPEATYSLILSCAKHGEALGLYD  EMKVSIVPSNYSVSLTYLYKNEYSKALSIFSEMEONKIVPDEVYIGLVYIK  LGIVEDAORMFEEDKAGLLSDQTVMAQAQVHMVQNTYDRLQVLDAMRANVPSQ  FYSALLRCHVAKEDVDAEDTRALSNYGPDPVFCNDLLRLYMLHLDKARALIL  MKREALQFDEDCVTVLECCCTINKDNLNTEQNEGSSKKNLPNDSTLSMM  KLKLLDPGGLSSVQLIMKFAREGSTDEAKFLYEHLTGAKPDPTATLIVQVQ  AQOLEQAQKLFTASTPVGVSYNVMDALCCGKTEAYRFLMELIDQGHNGAV  TISILYTLTKQEFQPAENIYVCLHDEALDTVYNTFKSMLEGLNDVLLVYS  YDRMISSGIPRSMOTENIMISYGGGKLEKAVEMSAQELGLPDKETYNMLSF  YKAGLHNEAIEIFQEKQNNHVPDSHTYLALTRATEGKCYSKAEALOMLRNWT  PSCTHNHLISAFLEKQIDEAQRMNOMEAGIPADLACCTMRMRHLDHGVDDGI  LFTFACRLKPDPSFILSAAFHLYHSGRESEAGVLDIAINMSGASFRLNKLKVGSLA  QIQPKWMSIVKSGPLFQAKAGVPLKYKQSPWCCSVEQVAREGEVDAQAQREGSD  AVHDHVORYDSSTCD"  complement(join(40023..40413,40774..40971,41292..41464,</p>	CDS
CDS	<p>complement(join(29217..30334,31321..31609,31701..31862))  /note="hypothetical protein"  /codon_start=1  /protein_id="BAA93020.1"  /db_xref="GI:7363276"  translation="MOSPSKSEASGAPATSVIEGMAELPEGLLHSIVALLGSLFD  LAFPTGCHSWRAAFSSYPSKSNFRILPPLVPRVVRKAPYSNNGHRLRSCVID  LNANRPLKCIQPEITQRHMFAGSHGQICCRGYCLVVDVFTGAEVSPPLPSE  NCDPEYGLITAPITSTOSHLIISTOSLDFWPGSDSWSLKLPRVNDVIFENG  OLIAVIEYSLYTLQAPILRLKIKTLMMDNNECPYMPFVYVCDMLLIVDHYISF  SFGAPVYRPLRDMTKPAKVEKKNLWALFIGDARSPFSCFNPRMGRGNC  LYAHYSQPSLSLHGLDADAVDNDNLVFKRWYRQLQALWVYPSMFYSAGDQ  complement(join(34717..34935,35401..35912,36242..36363,  36470..36542,36626..36754,36936..37247,37336..38617,  38724..39212))  /note="Similar to Arabidopsis thaliana DNA chromosome V,  BAC clone TM021B04; N. tabacum membrane-associated  salt-inducible protein. (AF007271)"  /codon_start=1  /protein_id="BAA93021.1"  /db_xref="GI:7363277"  translation="MAAIIAGAAATAVSTICSYSYEDDGAATSWLSGTSSTSSSS  SORORYRLLHDEACRLRRERGOGSAGTPRWRTTDOMARYVEDDRAGHVGRH  VVAVAAARATASCSPSSADMRANASFAKTIFREKCEVLHELHGWGRFPAW  KLQCEPVSVAITILLRLQGVQKVLAEVTFLEMLQACEPDVACGILLACAYAK  KGLNDMLFYAARRDIVPSISVFNFMSSQKQLHGVILHQBOMLEANYAPNOF  TYTVVIGSAKEGMELEADAFEMRRRREPEATYSLILSCAKHGEALGLYD  EMKVSIVPSNYSVSLTYLYKNEYSKALSIFSEMEONKIVPDEVYIGLVYIK  LGIVEDAORMFEEDKAGLLSDQTVMAQAQVHMVQNTYDRLQVLDAMRANVPSQ  FYSALLRCHVAKEDVDAEDTRALSNYGPDPVFCNDLLRLYMLHLDKARALIL  MKREALQFDEDCVTVLECCCTINKDNLNTEQNEGSSKKNLPNDSTLSMM  KLKLLDPGGLSSVQLIMKFAREGSTDEAKFLYEHLTGAKPDPTATLIVQVQ  AQOLEQAQKLFTASTPVGVSYNVMDALCCGKTEAYRFLMELIDQGHNGAV  TISILYTLTKQEFQPAENIYVCLHDEALDTVYNTFKSMLEGLNDVLLVYS  YDRMISSGIPRSMOTENIMISYGGGKLEKAVEMSAQELGLPDKETYNMLSF  YKAGLHNEAIEIFQEKQNNHVPDSHTYLALTRATEGKCYSKAEALOMLRNWT  PSCTHNHLISAFLEKQIDEAQRMNOMEAGIPADLACCTMRMRHLDHGVDDGI  LFTFACRLKPDPSFILSAAFHLYHSGRESEAGVLDIAINMSGASFRLNKLKVGSLA  QIQPKWMSIVKSGPLFQAKAGVPLKYKQSPWCCSVEQVAREGEVDAQAQREGSD  AVHDHVORYDSSTCD"  complement(join(40023..40413,40774..40971,41292..41464,</p>	CDS





```
* 43522 43621: gap of 100 bp
* 43622 45933: contig of 2312 bp in length
* 45934 46033: gap of 100 bp
* 46034 48737: contig of 2704 bp in length
* 48738 48837: gap of 100 bp
* 48838 51999: contig of 3062 bp in length
* 51900 51999: gap of 100 bp
* 52000 55449: contig of 3450 bp in length
* 55450 55549: gap of 100 bp
* 55550 60135: contig of 4586 bp in length
* 60136 60235: gap of 100 bp
* 60236 63645: contig of 3410 bp in length
* 63646 63745: gap of 100 bp
* 63746 66221: contig of 2476 bp in length
* 66222 66321: gap of 100 bp
* 66322 71307: contig of 4986 bp in length
* 71308 71407: gap of 100 bp
* 71408 75445: contig of 4038 bp in length
* 75446 75545: gap of 100 bp
* 75546 81775: contig of 6230 bp in length
* 81776 81875: gap of 100 bp
* 81876 87070: contig of 5195 bp in length
* 87071 87170: gap of 100 bp
* 87171 91815: contig of 4645 bp in length
* 91816 91915: gap of 100 bp
* 91916 98112: contig of 6197 bp in length
* 98113 98212: gap of 100 bp
* 98213 103688: contig of 5476 bp in length
* 103689 103788: gap of 100 bp
* 103789 111263: contig of 7475 bp in length
* 111264 111363: gap of 100 bp
* 111364 118151: contig of 6788 bp in length
* 118152 118251: gap of 100 bp
* 118252 124362: contig of 6111 bp in length
* 124363 124462: gap of 100 bp
* 124463 133078: contig of 8616 bp in length
* 133079 133178: gap of 100 bp
* 133179 139647: contig of 6469 bp in length
* 139648 139747: gap of 100 bp
* 139748 146970: contig of 7223 bp in length
* 146971 147070: gap of 100 bp
* 147071 156784: contig of 9714 bp in length
* 156785 156884: gap of 100 bp
* 156885 171574: contig of 14690 bp in length.
```

## FEATURES

## source

```
Location/Qualifiers
1..171574
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="17"
  /map="17"
  /clone_lib="RPC1-11 Human Male BAC"
  /clone="RP11-248L3"
1..1038
  /note="assembly_fragment"
misc_feature
1139..2373
  /note="assembly_fragment"
2474..3537
  /note="assembly_fragment"
3638..4776
  /note="assembly_fragment"
4877..6026
  /note="assembly_fragment"
6127..7205
  /note="assembly_fragment"
7306..8344
  /note="assembly_fragment"
8445..9543
  /note="assembly_fragment"
9644..10869
  /note="assembly_fragment"
10970..12118
  /note="assembly_fragment"
12219..13641
```

```
misc_feature
  /note="assembly_fragment"
13742..15584
  /note="assembly_fragment"
15685..17014
  /note="assembly_fragment"
17115..18904
  /note="assembly_fragment"
19005..20212
  /note="assembly_fragment"
20313..21330
  /note="assembly_fragment"
21431..22568
  /note="assembly_fragment"
22669..24914
  /note="assembly_fragment"
25015..26711
  /note="assembly_fragment"
26812..28657
  /note="assembly_fragment"
28758..30770
  /note="assembly_fragment"
30871..32848
  /note="assembly_fragment"
```

Query Match 3.1%; Score 67.8; DB 2; Length 171574;  
Best Local Similarity 46.3%; Pred. No. 0.11;  
Matches 216; Conservative 0; Mismatches 251; Indels 0; Gaps 0;

```
QY 6 caccggaactgtccgctttctggtttcttctgtgtgtgcgcctatgcctttccc 65
DB 20807 CCCCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 20866

QY 66 gactgcgcgcaatgctcttaacctttcaacctgcctcctcgtgtgcagctccagc 125
DB 20867 CCCCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 20926

QY 126 tcgctttgtctctgcagctcctcctcgtcgcgcacgaacacacagcgcctccg 185
DB 20927 CCCCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 20986

QY 186 tegtgcactcgtgtacgcgcctaccaggtacttgaacgagacgcgcgacttaact 245
DB 20987 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 21046

QY 246 gttgvcggaatccgtctacgctcgtcgtcagctcagctcctcagcgcgcgca 305
DB 21047 CGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 21106

QY 306 cgcacaaggcgtccgcgaacgcgactgagtgtgaccgcatctgttgccggttagcagg 365
DB 21107 CCCCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 21166

QY 366 gaaccaacacgacaagggttgcgcgcctagcaacagctcgcagcagcgcgcgca 425
DB 21167 GCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 21226

QY 426 aacaggcgtcggaggttgctcttcctcaatgtcgttgcctcccgcc 472
DB 21227 CGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 21273
```

## RESULT 15

```
LOCUS          GCU02524          1635 bp    DNA    linear    PLN 16-OCT-1993
DEFINITION     Geotrichum candidum NRRL Y-552 lipase gene, partial cds.
ACCESSION      U02524
VERSION        U02524.1   GI:408458
KEYWORDS
SOURCE
  ORGANISM      Galactomyces geotrichum.
                 Galactomyces geotrichum
                 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
                 Saccharomycetales; Dipodascaceae; Galactomyces.
REFERENCE      1 (sites)
```



AUTHORS	Bertolini, M., Laramee, L., Thomas, D., Cygler, M., Schrag, J. and Vernet, T.
TITLE	Polymorphism in the lipase genes of <i>Geotrichum candidum</i> strains
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 1635)
AUTHORS	Bertolini, M.
TITLE	Direct Submission
JOURNAL	Submitted (13-OCT-1993) Bertolini M., Biotechnology Research Institute - NRCC, 6100 Royalmount Avenue, Montreal, Quebec H4P 2R2, Canada

```

FEATURES             Location/Qualifiers
    1..1635
     /organism="Galactomyces geotrichum"
     /strain="NRRL Y-552"
     /db_xref="taxon:27317"
    <1..1635
     /codon_start=1
     /product="lipase"
     /protein_id="AAA03428.1"
     /db_xref="GI:408459"
     /translation="QAPTAVLNGNEVISGVLEKGVTFKGI PFADPVGVLRFKHPQNP
FWGSGVGLKANDFFSACMOLDPCNAISLDDKVVGLCKIIPDNLRGPLYDMAQGSVFPV
ECCILINVRPAGTRPDALKLVNVRWIIIGAFVFGSSASIPGNGIVKESVEMGQVFVVF
SINYTGPGYVFGGGDANTYAGKGLITQSGGGLPVSDNIANFGDDPKDVMIFGSGA
GAMVAHQIYVAGGDDNYAGKGLIFSAIQSGGGLPVFDSTSVGPSASVSRFAYQSGA
DASAGNETNACLRSKSDSLVISAQNSYDLKDLGLLPQFGPDPGNIIPDAAYE
LVRSRYAKVPIYTGNODEGHTILAPVATNATTPHVKKWLKYICSEASDAPLDRAVLS
LPGSWSEGAPERTGILNALTQFKRIAAIFTDLLQSPRRVYMLNATKQVNRRTYLAT
QAHNLLQIHTGSDLLQVFDLGPSSARRYFISFANHHDPPNVGTLNQDMQMDTYLAT
QKHEMLQIHMIGNSMRTDFFRIEGTISNFESDVTFLG"
326 a      449 c      391 g      469 t

```

Query Match	3.0%;	Score	67.6;	DB	8;	Length	1635;
Best Local Similarity	56.4%;	Pred. NO.	0.32;				
Matches	149;	Conservative	0;	Mismatches	109;	Indels	6;
Gaps							
1:							
QY	437	gaggattgcctcttcctcaatgcgttggcccccgcgcgtcgtcgaggggcgacaattctt	496				
Db	307	GAGGACTGTCTCACTTAACTTTCGGCCCGCTGGCACCAAGCCTGATGCTAAGCTT	366				
QY	497	cccgctcctgctcacattccagagtggtgcacgttcctggc-----gatgcgagacc	550				
Db	367	CCCGTCATGTTTGGATTACGGTGGTGGCCCTTTGTGTTGTTCTTCGCTCTTACCCCT	426				
QY	551	ggcagcgactttccgcgcttcacaaacacacggaacacgaatggctcggttataatctc	610				
Db	427	GSTAACGGCTACGTCGAAGGAGAGTGTGGAATGGGTGAGCCTGTTGTTGTTTCCATC	486				
QY	611	cagtcacgctcgcgagcttggttttcttcgcctgagccaaagccatgaaggactacgggta	670				
Db	487	AATACCGTACCGGCCCTATGGAATCTGGTGGTGATGCGCATACCGCTGAGGGTAAC	546				
QY	671	acgaacgcggcgcttgcttgaccag	694				
Db	547	ACCAACGCTGGTGTGCACGACCAG	570				

Search completed: June 29, 2002, 15:19:58  
Job time: 21479 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 29, 2002, 09:21:59 ; Search time 7060.77 seconds  
(without alignments)  
5085.839 Million cell updates/sec

Title: US-09-801-852A-1  
Perfect score: 1716  
Sequence: 1 atgtcttaacctttcac.....tctcggtagacgggtctc 1716

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues 3595312  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

# Database :

GenEmbl.\*  
1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*  
29: em.vi.\*  
30: em.htg.hum.\*  
31: em.htg.inv.\*  
32: em.htg.other.\*  
33: em.htgo.inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	ID	Description
-----					

not  
pind  
pind

1	1716	100.0	1716	6	AX268011
2	1716	100.0	1738	6	AR034144
3	1716	100.0	1738	6	BD005938
4	1714	99.9	1738	6	AF025410
5	1353	78.8	2220	6	AR034145
6	1353	78.8	2220	6	AX268013
7	1353	78.8	2220	6	BD005939
8	114.8	6.3	1855	8	CY002524
9	108.4	6.3	1635	8	CU002524
10	107	6.2	1635	8	A48376
11	107	6.2	1635	8	GCTAGL
12	107	6.2	1635	8	CU002387
13	107	6.2	1635	8	CU002341
14	106.8	6.2	1635	8	CU002622
15	106.8	6.2	1692	6	E02678
16	105.4	6.1	1635	8	CU002525
17	105.4	6.1	1635	8	CU002623
18	105.4	6.1	1635	8	CU002625
19	105.4	6.1	1674	6	E02497
20	105.4	6.1	1767	8	GCALIP2
21	105.4	6.1	1783	8	AB000260
22	105.4	6.1	1828	6	A74255
23	105.2	6.1	1812	8	CCLIP5
24	103.6	6.0	2043	8	CCLIP2
25	102.2	6.0	1657	8	CCLIPASE
26	102.2	6.0	1733	6	A48377
27	102.2	6.0	1733	6	CCLIP1
28	102	5.9	1647	6	A81171
29	96.8	5.6	1785	8	CCLIP4
30	91.4	5.3	1605	6	A81172
31	89.4	5.2	1688	12	AF044078
32	89.2	5.2	1692	8	GCCLR
33	88.8	5.2	45313	1	SCD95A
34	79.6	4.6	38209	1	SC5G9
35	78.6	4.6	1834	3	AF233526
36	78.6	4.6	1860	3	AY051959
37	74.6	4.3	24045	1	AB070941
38	73.4	4.3	2045	6	A45569
39	73.4	4.3	2045	6	A47607
40	73.4	4.3	2045	6	AR019333
41	68.4	4.0	1920	3	BMAJ3955
42	68	4.0	1927	10	AB010632
43	68	4.0	1965	10	BC015290
44	67.2	3.9	1611	6	AX121338
45	67.2	3.9	158063	9	AP001046

## ALIGNMENTS

RESULT	1	AX268011	Sequence 1 from Patent WO0166767.	1716 bp	DNA	linear	PAT 26-OCT-2001
LOCUS	AX268011	Sequence 1 from Patent WO0166767.					
DEFINITION	AX268011	AX268011.1 GI:16516554					
ACCESSION	AX268011	Rhodospiridium toruloides.					
VERSION	AX268011.1	Rhodospiridium toruloides					
KEYWORDS		Rhodospiridium toruloides					
SOURCE		Eukaryota; Fungi; Basidiomycota; Urediniomycetes; Microbotryomycetidae; Heterogastriales; Sporidiobolaceae; Rhodospiridium.					
ORGANISM		1 (sites)					
REFERENCE		Chiang, S.J. and Basch, J.D.					
AUTHORS		Direct production of deacetylcephalosporin c					
TITLE		Patent: WO 0166767-A 1 13-SEP-2001;					
JOURNAL		BRISTOL-MYERS SQUIBB COMPANY (US)					
FEATURES		Location/Qualifiers					
source		1..1716					
		/organism="Rhodospiridium toruloides"					
		/db_xref="taxon:5286"					
BASE COUNT		325 a	630 c	439 g	322 t		
ORIGIN							

Query Match 100.0%; Score 1716; DB 6; Length 1716;  
Best Local Similarity 100.0%; Pred. No. 3.6e-235;  
Matches 1716; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgtcttaacctcttcaacccctgcctccctcgtcgtcgcagctccagctcgcctttggc 60  
DB 1 ATGCTCTTAACCTCTTCAACCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCG 60  
QY 61 tctccagctcctcgtctccgc 120  
DB 61 TCTCCGACCTCCCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCG 120  
QY 121 ggtctacccgcctaccaggtctctgaagagaccgcgcgcgcgcgcgcgcgcgcgcgc 180  
DB 121 GGTCTACGCCCTACCAAGGCTCTTGAACGAGACCGCGGACCTCTCTCTCTCTCTCT 180  
QY 181 atccgctacgcctcggctcagcgtctccagctcctcagacccgcgcgcgcgcgcgcgc 240  
DB 181 ATCCGCTACGCTCGGCTCAGCGCTTCCAGGCTCTCTCAGACGCCCGCGACGACAA 240  
QY 241 gtcgcgaacgcgactgagatgagccgactgttggccggtcgtcgcgcgcgcgcgcgcgc 300  
DB 241 GTCCGCAACGGGACTGAGTATGGACCGGATCTGTGGCCGCTAGCGAGGGAACCAAC 300  
QY 301 accaaggcttgc 360  
DB 301 ACCAAGGCTTGGCCGCCCTTAGCAACAGCTCGAGACGCGCGCGCAACACAGCGCTG 360  
QY 361 gaggaattgcctctctcaatgtcgttcccccgcgcgcgcgcgcgcgcgcgcgcgcgcgc 420  
DB 361 GAGGATTGCCCTTCTCAATGTCTGTGGCCGCCGCGCTCGTGGAGGGGACAAATTT 420  
QY 421 cccgctcctcgtctacattcaaggagtggtctacgcttcggtcgtcgcgcgcgcgcgcgc 480  
DB 421 CCGCTCCCTCGCTACATTCACGGAGGTGGCTACGGCTTTCGGCGATGCGAGCACCGG 480  
QY 481 gacttgcgccttcaacagcacagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 540  
DB 481 GACTTTGGCGCTTCAACAGCACACGGGACCAAGATGGTCTGTGTAATCTCCAGTAC 540  
QY 541 cgtctcgcgcagcttggcttctcgtcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 600  
DB 541 CGTCTCGGACGCTTTGGTTCTCTCGCTGGCCAAAGCATGAAGGACTACGGGTAAAC 600  
QY 601 gcgcgcttgcgtcgc 660  
DB 601 GCGCGCTTGTGTGACACGCAATTCGCCCTTCAATGGGTTCAACACGACGCTCTCGA 660  
QY 720 ggcgcgaaccccgatcacggttacgatttggggcgagctcgcgcgcgcgcgcgcgcgcgcgc 720  
DB 661 GCGCGCAACCCCGATCACGTTACGATTGGGGCGAGCTGCGAGGCGCAGGGTCCGTTAT 720  
QY 721 aaccagatcattgcgaacggcgccacacccgtcgaagctcctcgttctcaagaagcccttc 780  
DB 721 AACAGATCATTTGCAACGGCGGCAACACCGTCAAGGCTCTCGGTCTCAAGAAGCCCTC 780  
QY 781 ttccacgctgcacgtcctcgtcctcctcctcctcctcctcctcctcctcctcctcctcct 840  
DB 781 TTCCACGCTGCCATCGGCTCTCCGCTTCTCCCTTCCCTACCAAGCAAGTACAACTCC 840  
QY 841 ttgcgcgagctcgtctactcccaactcgtctcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 900  
DB 841 TTGCGCGACGCTGCTACTTCCCACTCGTCTCGGGGACAACTGACACCAAGCCGCTCG 900  
QY 901 tcttcgcttgcctcgaagctcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 960  
DB 901 TCCTTCGCTTGCTCGAAGCTGTGACGCTGCGGGCTCGCTGCGCGCGGCGCTGAAGAAC 960  
QY 961 tcggcgcgcttcccgcttcgggttttggctgctatgtatcctcccgctcgtcgcgcgcgcgc 1020  
DB 961 TCGCGCGGCTTCGCTTCGGGTTTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1020

QY 1021 actgagcgcgctcgctcctctcgtcccaaggcgaagaggcgaagaaacctcaatggcaacctcttc 1080  
DB 1021 ACTGAGCGCGCTCGCTCCTCTCTCGCAAGGCAAGGAACCTCAATGCAACCTCTTC 1080  
QY 1081 accgggataaacaacctcgcagcgaaggattcattcactgacccactattcagaacgac 1140  
DB 1081 ACCGGGATCAACAACCTCGAGCGAGGATTCTATTTACTGACGCCACTATTTCAGAACGAC 1140  
QY 1141 acgatacgcagcagctcgcagcgcgtctccagttcgcgcgcgcgcgcgcgcgcgcgcgcgc 1200  
DB 1141 ACGATACGAGCCAGTCTCGAGCGCTCTCCAGTTCCAGCCGCTCTCGCGGCTCTTC 1200  
QY 1201 cccatacactcctcggaggc 1260  
DB 1201 CCTTACATCACTCGGAGGCGCGCAGCGCTCGCAAGCAGTACCCGATCTCCGAGCGG 1260  
QY 1261 ccgtcaaggcgaacacctctcgcactcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1320  
DB 1261 CCGTCAAGGGCAACCTTCTCTCGCATCTCGGCCGTCTATCGCGGACTCGACCTTCGTC 1320  
QY 1321 tgccgcactactgagccgc 1380  
DB 1321 TGCCCGACCTACTGAGCGCGGAGCGCTTGGCTCTCGGCCACAAAGGGCTCTTCGAC 1380  
QY 1381 tacgc 1440  
DB 1381 TACGCGCGGCTACCAACGCGACGACAACTCGTACTACATCGGCTCCATCTGGAACGGC 1440  
QY 1441 aagaagtcgctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 1500  
DB 1441 AAGAAGTCGCTCTGCTCGCTCCAGTCTCTTCAGCGCGGCTCGCGGGCTTCATCGAGAGG 1500  
QY 1501 ttcaaccgcaacaacacgcgtcgcgaagaccatcaaccccttactgcccgcgcgcgcgcgcgc 1560  
DB 1501 TTCAACCGCAACAACACGCTGCCAACAAGACCATCAACCTTACTGCGCCGACGTTGAC 1560  
QY 1561 tcgggcaagcagctcctcttcaacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1620  
DB 1561 TCGGCAACGAGCTCTCTTCAACGACGACGAGGACACCCCTCTCTCCCGCCGACCCG 1620  
QY 1621 cgcactggtgagacttcaagcttgc 1680  
DB 1621 CGCATGCTTGAGACTTCAAGCTTGACGACTTTGGCAGAGCCAGGACCAAGTGGCAG 1680  
QY 1681 ttctggtggtggtcaatcctcgtggaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1716  
DB 1681 TCTGGCGTGGTCAATCTCGGTGAACGCGGCTCTC 1716

RESULT 2  
AR034144 LOCUS AR034144 1738 bp DNA linear PAT 29-SEP-1999  
DEFINITION Sequence 1 from patent US 5869309.  
ACCESSION AR034144  
VERSION AR034144.1 GI:5949749  
KEYWORDS SOURCE Unknown.  
ORGANISM ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 1738)  
AUTHORS Politino,M., Tonzi,S.M., Usher,J.J., Burnett,W.V. and Romancik,G.  
TITLE Cephalosporin esterase gene from Rhodosporidium toruloides  
JOURNAL Patent: US 5869309-A 1 09-FEB-1999;  
FEATURES Location/Qualifiers  
1. 1738  
SOURCE /organism="unknown"  
BASE COUNT 327 a 639 c 444 g 328 t  
ORIGIN  
Query Match 100.0%; Score 1716; DB 6; Length 1738;  
Best Local Similarity 100.0%; Pred. No. 3.6e-235;



FEATURES		Location/Qualifiers	
source	1..1738	/organism="unidentified"	
BASE COUNT		327 a	639 c 444 g 328 t
ORIGIN			
Query Match 100.0%; Score 1716; DB 6; Length 1738;			
Best Local Similarity 100.0%; Pred. No. 3.6e-235;			
Matches 1716; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	atgtccttaacctctcaacctcgctccctcgctgagcgtccagctcgctttggcc	60
Db	11	ATGCTCCTTAAGCTTTCACGCTCGCTCCCTCGCTGCGAGCGCTCCAGCTCGCTTTGCC	70
Qy	61	tctcgaacctccctcgctcgccgcaacgacccaaagagcccccctccgctcgacatc	120
Db	71	TCCTCGACCTCCCTCGCTCGCGCGACGAAACCAACGAGCCCTCCCTCGCTCGACCTC	130
Qy	121	ggctacgcccgtaccaggtacttgaacgagaccgcgagctactctgtggcgcgga	180
Db	131	GGCTACGCCGCTACCAAGGCTACTTTGAACGAGACCGCGGACTACTGTGTGGCGCGA	190
Qy	181	atcgcgtacgctcggtcgagcgtctccaggtcctcagacgcccgcgcgcacaaagcc	240
Db	191	ATCCGCTACGCTCGGCTCAGCGCTTCCAGGCTCCTCAGACGCCGCGACGACACAAGGCC	250
Qy	241	gtccgaacgcgactgagtatgacgactctgttggccggttagcaggggaacaaacag	300
Db	251	GTCCGCAACCGGACTGAGTATGACCGATCTGTTGGCCGGCTAGCGAGGGAACCAACAGC	310
Qy	301	accaagggctgcgcgccttagcaacagctcgagcgcgcgcgcgcgcgcgcgcgcgcgc	360
Db	311	ACCAAGGCTTGGCGCGCTTACACAGCTGAGCAGCAGCGCGCGCGAGAAACAGCGCTCG	370
Qy	361	gaggattgcctcttctcaatgtctgttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	420
Db	371	GAGGATTGCTCTTCTCAATGTCTGTTGCCCGCGCGCTCGTGGCGAGGGCGACAATCTT	430
Qy	421	ccgctcgtctacattcaagaggtggttaagccttcggcgatcgagcagccggagc	480
Db	431	CCGCTCTCTGTACATTACGAGAGTGGCTACGCGCTTCGGCGATGCGAGCACCGGCGAC	490
Qy	481	gacttgcgccttcaacagcacgcgggaacaaagatggtgttgaattctccagttac	540
Db	491	GACTTTGCGGCTTCAACAGCACACGGGACCAAGATGGTGTGTTAAATCTCCAGTAC	550
Qy	541	cgctcggcagcttgggttctcgtcgcccaagccatgaagactacggtgtaacgaac	600
Db	551	CGTCTCGGAGCTTGGTTTCTCGCTGCGCAAGCCATGAAGGACTACGGTGTAAACGAAC	610
Qy	601	gcggctgttgacagaacatcgcccttcaatgggttcaacagcagctctcgaagttc	660
Db	611	GCCGGCTGTGTGACCAAGCAATTCGCCCTTCAATGGGTTCAACAGACGCTCTCGAAGTTC	670
Qy	661	ggcgcaaccccgatcaagtttggggcgagctcgcggcgaggggtccggttatg	720
Db	671	GCGCGCAACCCGATCACGTTACGATTGGGGCGAGTCTGACGGCGCAGGGTCCGTTATG	730
Qy	721	aaccagatcattcgaaagcggcgaacacgcgtcaaggtcctcggttcaagaagcccttc	780
Db	731	AACCAAGATCATTCGGAACGGCGCAACACCGTCAAGGCTCTCGGTCTCAAGAAGCCCTTC	790
Qy	781	ttcaagctgcatcggtcctcgttcttctccctacacgaacgaagtaacactcccc	840
Db	791	TTCCAGCTGCCATCGGCTCTCTCGCTTCTCTCCCTACCAAGCAAGTACAACCTCCCTCC	850
Qy	841	ttcgcgagctgcttactcccaactcgtctcggcgcaaaactgcacaaagccgctcg	900
Db	851	TTCCGCGAGCTGCTTACTCCCAACTGCTCTCGGCGACAAACTGCAACCAAGCCGCTCG	910
Qy	901	tctctgcttgcctcgaagctgtcgacgctgagggcgtcgtcgtcgagcgagcgtgaagac	960
Db	911	TCCTTCGCTTGGCTCGAAGCTGTGACGCTCGCGCGCTCGTGTGCGCGGGGCTGAAGAAC	970
Qy	961	tcggcgcggttcccggttcgggttttgggtcgatgtcccggttcgctcgacgagcctctctg	1020
Db	971	TCGGCGGGGTTCCCGTTCGGGTTTGGTGTATGTGCCGGTGTCTGACGGGACCTTCTTG	1030
Qy	1021	actgagcgcgctcgctcctctcctccaaagggaagaacacctcaatggcaacctcttc	1080
Db	1031	ACTGAGCGCGCTCGCTCCTTCTTCGCAAGGGCAAGAAAGACCTCAATGGCAACCTCTTC	1090
Qy	1081	accgggatcaacaacctcgacgaaggtatcatatctacatgacgacacatattcagaacac	1140
Db	1091	ACCGGATCAACAACCTCGACGAAGGATTATATTACTGTAGCGGCACATTATCAGAAGCAC	1150
Qy	1141	acgacagcagcagctcgagcgctctcccggttcgacgacgctcctcgcggtctcttc	1200
Db	1151	ACGATCAGGACGATCGACGCGCTCTCCAGTTCCGACCGCTCTCTCGCGGCTCTTTC	1210
Qy	1201	ccctacatcacctcgaggagcgcagcgcgtcgcggaagcagtagtaccgatctccgaagcg	1260
Db	1211	CCCTACATCACTCGGAGGAGCGCAGGCGCTCGGAAAGCAGTACCCGATATCTCGACGCG	1270
Qy	1261	cgtcaaaaggcaaaccttctctcgcatctcgccgctcatcgaggatcgacactctctc	1320
Db	1271	CCGTCAAGGGCAACACCTTCTTCGCATCTCGCGCGCTCATCGCGGACTCGACCTTCGTC	1330
Qy	1321	tgcccgactactgagcccgagcgttcgctcgtcgcgcgcgcgcgcgcgcgcgcgcgcgc	1380
Db	1331	TGCCCGACCTACTTGACCGCGCGGAGGCTTGGCTGTCTGCGGCCCAAGGGGCTCTTCGAC	1390
Qy	1381	tacgcgcggtccaccacgc	1440
Db	1391	TACGCGCGGCTCACACGCGACCGGAACTCGTACTACTATCGGCTCCATCTGGAACGCG	1450
Qy	1441	aagaatcggttcgctcgtccagtccttcgacgcgcgctcggcggttccatcgagacg	1500
Db	1451	AAGAAGTCGCTTCGCTCGCTCAGTCTTCGACGCGCGCTCGCGGGCTTTCATCGAGACG	1510
Qy	1501	ttaacccgcaacaacacgcgtgccaaagacacatacccttacttgccgacgcttcgac	1560
Db	1511	TTCAACCCGANCAACACGCTGCCAACAAAGACATCAACCTTACTTGGCGACGTTTCGAC	1570
Qy	1561	tcgggaagcagctcctcttcaacacgacgagcagggacacccctctctccgcgcacccg	1620
Db	1571	TCGGCAAGCAGCTCCTCTTCAACAGCAGCAGGAGGACACCTCTCTCCCGCGCACCG	1630
Qy	1621	cgcctgttgagacttcaagcttgaccgactttgacgagcagcagcagcagcagcagcagc	1680
Db	1631	CGCATCTGTGAGACTTCAAGCTTGACCGACTTTGGCAGGCGCAGAGCAAGTGGCGAC	1690
Qy	1681	tctcgcggtgggtcaatctcgtggaacgcggtctc	1716
Db	1691	TTCTGGCGTGGGTCAATCTCGGTGAACCGGGTCTC	1726
RESULT 4			
AF025410 1738 bp mRNA linear PLN 30-DEC-1997			
LOCUS Rhodospiridium toruloides cephalosporin esterase precursor, mRNA,			
DEFINITION complete cds.			
ACCESSION AF025410			
VERSION AF025410.1 GI:2731567			
KEYWORDS Rhodospiridium toruloides.			
SOURCE Rhodospiridium toruloides.			
ORGANISM Rhodospiridium toruloides			
REFERENCE Eukaryota; Fungi; Basidiomycota; Uredinomyces;			
AUTHORS Microbotryomycetidae; Heterogastridiales; Sporidiobolaceae;			
TITLE 1 (bases 1 to 1738)			
Polittino,M., Tonzi,S.M., Burnett,W.V., Romancik,G. and Usher,J.J.			
Purification and characterization of a cephalosporin esterase from			
Rhodospiridium toruloides			







Db 1757 GGGCTCTTCAGTACGCGCGGCTCACCAGCGACGACCACTCGTACTACATCGGGTC 1816  
Qy 1428 catctggaacggaagagtgctcgtccgtccagtccttcgagcgcgctcgcg 1487  
Db 1817 CATCTGGAACGGCAAGAGTGGTCTCGTCCGTCAGTCCCTTCGACGGCGGCTCGGGCG 1876  
Qy 1498 ctteatcgagagcttaacccggaacaaacgctgcccgaacagaccattactg 1547  
Db 1877 CTTTCATCAGAGCTTCAACCCGGAACACAGCGTGCACCAAGACCATCAACCCCTTACTG 1936  
Qy 1548 gccagcttcaactcgggcaagagctcctcttcaacagcagcagaggggacacctc 1607  
Db 1937 GCCGACGTTGACTCGGGCAAGAGCTCCTCTTCAACAGCAGCAGAGGGACACCCCTC 1996  
Qy 1608 tccgcgacccggcagctcgttgagacttcaagcttgcagacttggcagcagcagaa 1667  
Db 1997 TCCGCGACCGCGGCATCGTTGAGACTTCAAGCTTGACCGACTTGGCAGAGGCCAGAA 2056  
Qy 1668 gaccaagtgcactcttgctgggtggtcaatctcgtgtaacggtgtc 1716  
Db 2057 GACCAAGTGCAGCTTCTGGCGTGGGTCAATCTCGGTGAACGGGTCTC 2105

RESULT 6  
AX268013  
LOCUS AX268013 2220 bp DNA linear PAT 26-OCT-2001  
DEFINITION Sequence 3 from Patent WO0166767.  
ACCESSION AX268013  
VERSION AX268013.1 GI:16516555  
KEYWORDS Rhodospiridium toruloides.  
SOURCE Rhodospiridium toruloides.  
ORGANISM Rhodospiridium toruloides  
Eukaryota; Fungi; Basidiomycota; Urediniomycetes;  
Microbotryomycetidae; Heterogastridiales; Sporidiobolaceae;  
Rhodospiridium.

REFERENCE  
AUTHORS Chiang, S.-J. and Basch, J. D.  
TITLE Direct production of deacetylcephalosporin c  
JOURNAL Patent: WO 0166767-A 3 13-SEP-2001;  
BRISTOL-MYERS SQUIBB COMPANY (US)  
FEATURES  
source Location/Qualifiers  
1..2220 /organism="Rhodospiridium toruloides"  
/db\_xref="taxon:5286"  
BASE COUNT 389 a 814 c 554 g 463 t  
ORIGIN

Query Match 78.8%; Score 1353; DB 6; Length 2220;  
Best Local Similarity 84.6%; Pred. No. 1.5e-183;  
Matches 1716; Conservative 0; Mismatches 0; Indels 313; Gaps 5;

Qy 1 atgctccttaaccttcaacctgctcctcgtcgtgagagctccagctcgtccttgc 60  
Db 77 ATGCTCCTTAACCTCTTCACTCCTCCCTCCCTCGTGGAGCTCCAGCTCGCTTGC 136  
Qy 61 tctccgacctcctcgtcgcgcagcaacccaaacagagccctcccgctgtagactc 120  
Db 137 TCTCGACCTCCCTGTCGCGGCAGCAACCCAAACAGAGCCCTCCGCTCGTCACTC 196  
Qy 121 ggctacgcccgtacaaagcttaactgaacagacccgagacttactggtggtgag 180  
Db 197 GGCTACGCCCGCTACCAAGGCTACTTGAACAGAGACCGCGGACTCTACTGGTGGCGGA 256  
Qy 181 atcgcctacgctggtcaggtcttcaggtcctcagagcccgccgacgcaagcc 240  
Db 257 ATCCGCTACGGCTCGGGCTCAGCGCTTCAGGCTCCTCAAGCGCCCGCAGCACAGGCC 316  
Qy 241 gtccgcaacgagctgagatgagccgactctgttgccggtctagcgagggaaccaacg 300  
Db 317 GTCCGCAACGGGACTGAGTATGGACCGATCTGTGGCCGGCTAGCGAGGGAAACACAG 376

Qy 301 accaagggcttgccgcttagcaacagctcagagcagcgccgcaaaaacagcgctcg 360  
Db 377 ACCAAGGGCTTGCCTGCGCGCTAGCAACAGCTCGAGCAGCGCGCGCAACAGCGCTG 436  
Qy 361 gagatgctctctcctcaatgctggtcccccgcggtcgtcgagagggcgacaattt 420  
Db 437 GAGATGCTCTCTCTCAATGCTGTGCCCGCGCGCTCGTGAGAGGGGACATCTT 496  
Qy 421 cccgtcctcgtctacattcacggaggtggtctacgcttcctcggcgtgagcagcgagc 480  
Db 497 CCGGTCTCTCTACATTCACGGAGGTGGCTACGCTTCGGCGATGTCGAGCAGCGCAGC 556  
Qy 481 gacttgcgctcctcaacaaacagggaaacaaagatggtcgttgaatactccagttac 540  
Db 557 GACTTGGCGGCTTCACCAAGCACACGGGAACCAAGATGGTGTGTAATCTCCAGTAC 616  
Qy 541 cgtctcgagcttctggttctcgtcgccaaagcacaagagactcaggtgtaacgaac 600  
Db 617 CGTCTCGGACACTTGGTTCCTCGTGGCCAAAGCATGAAGGACTACGGTGTACGAAC 676  
Qy 601 gccggttgccttgac----- 615  
Db 677 GCCGGCTTGTGACCAAGGTGAGTTTCCCGCATGATACCCGCCACCTTTCGACTCATGC 736  
Qy 616 -----cagcaattcgccttcaatgggttcaacagcagctctcgaag 657  
Db 737 TGACGCCCTCTCCCGCTCGCAGCAATTCGCCCTTCAATGGTTCAACAGCAGCTCTCGAAG 796  
Qy 658 ttcggcgcaaaccccgatcacgttaccgatttggggcgagctctcagcgcgaggtccgtt 717  
Db 797 TTCGGCGCAACCCCGATCACGTTACGATTTGGGGCGAGTCTGCAAGGCGCAGGTTCCGTT 856  
Qy 718 atgaaccagatcattgcgaac----- 738  
Db 857 ATGAACAGATCATTTGCGAAGCTGAGCCACCGCAACCGATCTCCAGCGGACTTTCCCCCC 916  
Qy 739 -----ggcgcaacacgcgtcaaggtctcgtg 764  
Db 917 CCCCCCCCCCGCTGACCTCCTCGTCTTGCAGGGCGCGCAACACCGCTCAAGGCTCTCGG 976  
Qy 765 tctcaagaagccctcttccagctgcgcgtccgtccgtcttccctccctccctacagc 824  
Db 977 TCTCAAGAAGCCCTCTTCCAGCTGCGATGGCTCTCCGCTCTCTCCCTCCCTACCAAGC 1036  
Qy 825 caagtacaactcctcctcgcgagctgctctactcctcctcctcctcctcctcctcctcct 884  
Db 1037 CAAGTACAACCTCCCGCTTTCGCGAGCTGCTCTACTCCCACTCGTCTCGGCGACAATG 1096  
Qy 885 cacaagaagcgcctcgtcctcgtcgtcgaagctgagcgtgagcgtcgtcgtcgtcgtcgtc 944  
Db 1097 CACCAAAAGCGCTCGTCTTTCGCTTGGCTCGAAGCTGTCGACGCTCGGCGCTCGCTGC 1156  
Qy 945 ggcggcgtgaagaactcgcgcgtcccgctcgggttttgggtggtatggtcccggtcgt 1004  
Db 1157 GCGGGCGGTGAAGAACTCGGGCGGCTTCCCGTTCGGGTTTGGTGTATGTCGCCGCTGT 1216  
Qy 1005 cgacgggaccttcttgcagtcgagcgctcgtcctctcgcgaagggaagaagaaact 1064  
Db 1217 CGACGGGACCTTCTTGAAGTGAAGCGCGCTCGCTCTCTTCGCAAGGGCAAGAGAAGCT 1276  
Qy 1065 caatggc----- 1071  
Db 1277 CAATGGCGTGGCTGGCGAGCTTTCGAGTCTTCAGGATCTCGCTGACACTGTCGACCGGC 1336  
Qy 1072 -----aaccttccacgggatacaaacctccgacgaag----- 1105  
Db 1337 TCGCAGAACCTCTTCCCGGGATCAACAACTTCGAGAGAGATGAGTTCCTCGTCGAGGCT 1396  
Qy 1106 -----gattcatattcactgac 1122  
Db 1397 CTGTTGCCCGCAGCGAGACTGACTTGTCTCTTTTTCGGAAGATTACGATTATATTTACTGAC 1456  
Qy 1123 gccactattcagaacgacagatcagcagctcagcagcagctcagcagcagctcagcagcagc 1182





421  
5y

100

Qy 421 cccgtcctcgtctacattcacggagggtggctacgccttcggcgatgcgagcaccggcagc 480

```
Db 471 CCGGTATGCTCTGGATCTTTGGCGGTGGGTTTGAGATCGCAGCCCCACCACCTTCCCT 530
Qy 481 gactttgcccgttcaccaa-----gcacacgggaacaaagatggtcggttgtaaatctc 534
Db 531 CCCGCCAGATGGTCAACAGAGTGTGCTATGGCAAGCCCATCATCCAGTGGCGGTC 590
Qy 535 cagtaccgtctcgcagcttggcttctcgtcgtgccaagcagcaggaactacaggtgta 594
Db 591 AACATACCGTGTCTCGTGGGGTCTTGCTGGTGTGATGACATCAAGCGCGAGGGGAGC 650
Qy 595 acgaacgcggcttcttgaccagcaaatcgcccttcaatgggttcaacagcagctctcg 654
Db 651 GGGAAACCGCGCTTGAAGGACACCGTTTGGCATGCGTGGTGGCAGACAAATGCCC 710
Qy 655 aagtcggcggaaccccgatcacgttacgatttggggcgagtcgagcgcgaggtcc 714
Db 711 GGGTTGGCGCGACCCGACGAGGTGACGATCTTTGGCGAGCTGGCGGGCAGCATGTC 770
Qy 715 gttatgaaccagatcattgcaacggcggaacacc 750
Db 771 GTGTGTGCCACCTCATCTGGAACGACGGCGACAAC 806

RESULT 9
GCU02524 GCU02524 1635 bp DNA linear PLN 16-OCT-1993
DEFINITION Geotrichum candidum NRRL Y-552 lipase gene, partial cds.
ACCESSION U02524
VERSION U02524.1 GI:408458
KEYWORDS
SOURCE Galactomyces geotrichum.
ORGANISM Galactomyces geotrichum
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
REFERENCE 1 (sites) Saccharomycetales; Dipodascaceae; Galactomyces.
AUTHORS Bertolini, M., Laramée, L., Thomas, D., Cygler, M., Schrag, J. and
Vernet, T.
TITLE Polymorphism in the lipase genes of Geotrichum candidum strains
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1635) Bertolini, M.
AUTHORS Bertolini, M.
TITLE Direct Submission
JOURNAL Submitted (13-OCT-1993) Bertolini M., Biotechnology Research
Institute - NRCC, 6100 Royalmount Avenue, Montreal, Quebec H4P 2R2,
Canada
FEATURES
source Location/Qualifiers
1. .1635
/organism="Galactomyces geotrichum"
/strain="NRRL Y-552"
/db_xref="taxon:27317"
<1. .1635
/codon_start=1
/product="lipase"
/protein_id="AA03428.1"
/db_xref="GI:408459"
/translation="QAPTAVLNGNEVLSGVLEKVDTFKIPFADPPVGLDRFKHPQP
FTGSGYGLKADFSSACMQDLPNGAISLKDVKVGLKIIIPDNLGRPLYDMAQGSVMN
EDCLYNVFPAGTKPAKLPVMWYIGAFVFGASYPNGYVXKESVEMGQPVFV
SINRTGYPVGLGDAITAEGNTNAGLDKORGLWVSDNIANFGGDPDKVMIFGESA
GANVAHQLVAYGDDNTYNGKQLFHSALQSGGLPFDSTSVGPSAYSRSFQAQAGC
DASAGNETLACRSKSSDVLHSAQSYDUDLGLLPQFLGFPDPDGNIIIPDAAYE
LYPSRGYAPFTITGNOEDGTILAPVAINATTPHVKWLYICSEASDAPLDRLYS
LYPSGSEKAPFTITGALNATLPQFKRIAAITFDLLFQSPRVMLNATKDVNRWTYLAT
QLHNLVPLGTFHSGDLLFOYVDLGLPSAYRRYFISFANHHPNVGTNLKQMDWTD
AGREMLQIHMIGNSMRTDFRIESINFEESDVILFG"
BASE COUNT 326 a 449 c 391 g 469 t
ORIGIN

Query Match 6.3%; Score 108.4; DB 8; Length 1635;
Best Local Similarity 56.6%; Pred. No. 1.9e-06;
Matches 224; Conservative 0; Mismatches 166; Indels 6; Gaps 1;
```

```
Qy 361 gaggattgctctcttcctcaatgctgttgccccgcggctcgctgagggcgacaaatctt 420
Db 307 GAGGACTGTCTCTACCTTAACGTTTTCCGCCCGCTGGCACCAAGCCTGATGCTTAAGCTT 366
Qy 421 cccgtctcgtctacattcacggaggtggctacgcttcggc-----gatggagaccc 474
Db 367 CCCGTATGTTGGATTACGGTGGTGCCTTTGTTGTTGGTTCCTTCTTACCTT 426
Qy 475 ggcagcgactttgcgccttcaccaagcacacgcgggaaccaagatggttcgttgtaaatctc 534
Db 427 GGTAAACGCCTACGTCAAGGAGAGTGTGGAATGGGTGAGCTGTGTGTTGTTTCCATC 486
Qy 535 cagtaacgcttcggcagcttggcttctcgtcgtgccaagccatgaaggactacgggtgta 594
Db 487 AACTACCGTACCGGCCCTATGGATTCTGGTGGTGTGATGCCATFCACCGTGAAGGTAAC 546
Qy 595 acgaacgcggcttcttgaccagcaattcgcccttcaatgggttcaacagcagctctcg 654
Db 547 ACCAACCGTGGTCTGCAGCACCGACGCAAGGTTCTCGATGGGTTAGCGACAACTGGCC 606
Qy 655 aagtcggcggaaccccgatcacgttacgatttggggcgagtcgagcgcgaggtcc 714
Db 607 AACTTTGGTGGTATCCCGACAAAGTCAATGATTTTCGTGAGTCCGCTGGTGCATGAGT 666
Qy 715 gttatgaaccagatcattgcaacggcggaacacc 750
Db 667 GTTGCTACCAGCTTGTGCTATGTTGGTGGTGAACAAC 702

RESULT 10
A48376 A48376 1635 bp DNA linear PAT 07-MAR-1997
LOCUS
DEFINITION Sequence 5 from Patent WO9603511.
ACCESSION A48376
VERSION A48376.1 GI:2302167
KEYWORDS
SOURCE Galactomyces geotrichum.
ORGANISM Galactomyces geotrichum
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
REFERENCE 1 (bases 1 to 1635) Saccharomycetales; Dipodascaceae; Galactomyces.
AUTHORS Alibert, G., Mouloungui, Z. and Boudet, A.
TITLE METHOD FOR PRODUCING FATTY ACIDS OR DERIVATIVES THEREOF FROM OIL
JOURNAL PLANT
Patent: WO 9603511-A 5 08-FEB-1996;
TOULOUSE INST NAT POLYTECH (FR)
Other publication AU 2984995 960222
Other publication FR 2722798 960126.
FEATURES
source Location/Qualifiers
1. .1635
/organism="Galactomyces geotrichum"
/db_xref="taxon:27317"
BASE COUNT 323 a 449 c 394 g 469 t
ORIGIN

Query Match 6.2%; Score 107; DB 6; Length 1635;
Best Local Similarity 56.1%; Pred. No. 2.9e-06;
Matches 225; Conservative 0; Mismatches 170; Indels 6; Gaps 1;
```

```

Db 487 AACTACCGTACCAGCCCTATGATCTCTGGGTGGTATGCCATCACCGCTGAGGGTAAC 546
      ||||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 595 acgaagccgctgtgtgaccagcattcgccctcaatgggttcaacagcagctctcg 654
      ||||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 547 ACCAAGCTGTCTGCAGCAGCAGCAGCAGGCTCTGAGTGGTGTAGCGACAACATTGCC 606
      ||||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 655 aagttcggcggaaccccgatcgttaacgatttggggcgagctctcgagcgagggctcc 714
      ||||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 607 AACTTTGGTGTATCCGACAAGGTCATGATTTTCGGTGAGTCCGCTGGTGCATGAGT 666
      ||||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 715 gttatgaaccagatcattcgaaacgagcggaacacacgtcaa 755
      ||||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 667 GTTGCTACCAAGCTTGTGCTACGGTGGTGACACACCTA 707
      ||||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

RESULT 11
GCTAGL GCTAGL 1635 bp DNA linear PLN 25-JAN-1996
LOCUS G.candidum triacylglycerol lipase gene.
DEFINITION X81656
ACCESSION X81656
VERSION 1
KEYWORDS triacylglycerol lipase.
SOURCE Galactomyces geotrichum.
ORGANISM Galactomyces geotrichum
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
REFERENCE 1 (bases 1 to 1635)
AUTHORS Pretorius, G.H.J.
TITLE Direct Submission
JOURNAL Submitted (16-SEP-1994) G.H.J. Pretorius, University of the Orange
Free States, Dept of Microbiology and Biochemistry, PO Box 339,
Bloemfontein, 9300, SOUTH AFRICA
REFERENCE 2 (bases 1 to 1635)
AUTHORS Phillips, A., Pretorius, G.H. and van Rensburg, H.G.
TITLE Molecular characterization of a Galactomyces geotrichum lipase,
another member of the cholinesterase/lipase family
JOURNAL Biochim. Biophys. Acta 1252 (2), 305-311 (1995)
MEDLINE 96049515
FEATURES
source Location/Qualifiers
1. .1635
/organism="Galactomyces geotrichum"
/strain="CBS 178.71"
/db_xref="taxon:27317"
<1. .1635
/EC_number="3.1.1.3"
/codon_start=1
/product="triacylglycerol lipase"
/protein_id="CAA57316.1"
/db_xref="GI:547504"
/translation="QAPTAVLNGNEIVSGVLEKGVDTFKGIPFADPPVGLRKFHPQP
FTGSYGLKANDFSACMOLDPNAISLDKVVGLGKILPDLNLRGLYDMAOGTSMN
ECLLYNVRPAGTKPDALPVWVLYGGAFFGSSASYPGNGYVKESVEMGPPVVF
SINRTGPGFLGGDAITAEAGNTNGLHQRKLEWNSDNIANFGDPPKVMIFGESA
GAMSAVHQLVAGDNTYNGKLFHSAIQNSYDLKDLFLGLPQFLGFGPRPDGNIIPDAAYD
DPSVADNETLACLRSSDYLHSAQNSYDLKDLFLGLPQFLGFGPRPDGNIIPDAAYD
LYRSGRYAKVPYITGNOEDGTILAPVAINATTTPHVKKWLYICSOASVSLDRVLS
LYPGSWSEGAFFRTGLNALTPOFKRIAAFTDQLLFQSPRRVMLNATKDVNRWYTLAT
QLHNLVFFLGTTHGSDLLFQYVVDLQSPSAYRYFFISFANHDPNVGTLNQWDMYTD
AGKMLQIHMVGNMRTDDFRIGISNFESDVLFG"
BASE COUNT 328 a 453 c 387 g 467 t
ORIGIN

Query Match 6.2%; Score 107; DB 8; Length 1635;
Best Local Similarity 56.1%; Pred. No. 2.9e-06;
Matches 225; Conservative 0; Mismatches 170; Indels 6; Gaps 1;

QY 361 gaggattgctctcctcaatgctgtgcccccgccgctgctgagcgagcacaatttt 420
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 307 GAGGACTGTCTCTACCTACAGTTTCCGCCCGCTGGCACCAGCCTGGAGATAAGCTC 366
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

QY 421 ccggtctctgtacattcaaggaggtggtacagccttcggc-----gatcgagcacc 474
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 367 CCCTCATGTTTGGATCTACGGTGGTGCCTTGTGTTGTTCTTCTGCTTCTTACCT 426
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 475 ggagcgacttgcgccttcaccaagcacacagcggaacacgaagatggctgtgtaaatctc 534
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 427 GGTAAAGGCTACGCTACAGAGAGAGTGTGAAATGGCCGCCAGTTGTGTTTTCATC 486
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 535 cagtacgctcgcgcagcttgggttctcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 594
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 487 AACTACCGTACCAGTCCCTATGATTTTCGGTGGTGTATGCCATCACCGCTGAGGGCAAC 546
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 595 acgaagccgctgtgtgaccagcattcgccctcaatgggttcaacagcagcgtctcg 654
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 547 ACTAAGCTCTGCTCTGCTACGACGACGAGGCTCTCAGTGGGTACGACCAACATTGCC 606
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 655 aagttcggcggaaccccgatcgttcaacgatttggggcgagctctcgagcgagggctcc 714
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 607 AACTTTGGTGTATCCGACAAGGTCATGATTTTCGGTGGTGGTGCCTGAGTGCATGAGT 666
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 715 gttatgaaccagatcattcgaaacgagcggaacacacgtcaa 755
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 667 GTTGCTACCAAGCTTATGCTACGGTGGTGACACACCTA 707
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 12
GCU02387 GCU02387 1635 bp DNA linear PLN 08-OCT-1993
LOCUS Geotrichum candidum NRRL Y-553 lipase gene, partial cds.
DEFINITION U02387
ACCESSION U02387
VERSION U02387.1
KEYWORDS GI:406507
SOURCE Galactomyces geotrichum.
ORGANISM Galactomyces geotrichum
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
REFERENCE 1 (sites)
AUTHORS Bertolini, M., Laramée, L., Thomas, D., Cygler, M., Schrag, J. and
Vernet, T.
TITLE Polymorphism in the lipase genes of Geotrichum candidum strains
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1635)
AUTHORS Bertolini, M.
TITLE Direct Submission
JOURNAL Submitted (05-OCT-1993) Bertolini M., Biotechnology Research
Institute - NRCC, 6100 Royalmount Avenue, Montreal, Quebec H4P 2R2,
Canada
FEATURES
source Location/Qualifiers
1. .1635
/organism="Galactomyces geotrichum"
/strain="NRRL Y-553"
/db_xref="taxon:27317"
<1. .1635
/codon_start=1
/product="lipase"
/protein_id="AAA03425.1"
/db_xref="GI:406508"
/translation="QAPTAVLNGNEIVSGVLEKGVDTFKGIPFADPPVGLRKFHPQP
FTGSYGLKANDFSACMOLDPNAISLDKVVGLGKILPDLNLRGLYDMAOGTSMN
ECLLYNVRPAGTKPDALPVWVLYGGAFFGSSASYPGNGYVKESVEMGPPVVF
SINRTGPGFLGGDAITAEAGNTNGLHQRKLEWNSDNIANFGDPPKVMIFGESA
GAMSAVHQLVAGDNTYNGKLFHSAIQNSYDLKDLFLGLPQFLGFGPRPDGNIIPDAAYE
LYRSGRYAKVPYITGNOEDGTILAPVAINATTTPHVKKWLYICSDASLDRLVLS
LYPGSWSEGAFFRTGLNALTPOFKRIAAFTDQLLFQSPRRVMLNATKDVNRWYTLAT
QLHNLVFFLGTTHGSDLLFQYVVDLQSPSAYRYFFISFANHDPNVGTLNQWDMYTD
AGKMLQIHMVGNMRTDDFRIGISNFESDVLFG"
BASE COUNT 323 a 449 c 394 g 469 t
ORIGIN

Query Match 6.2%; Score 107; DB 8; Length 1635;
Best Local Similarity 56.1%; Pred. No. 2.9e-06;

```

Matches	225; Conservative	0; Mismatches	170; Indels	6; Gaps	1;
QY	361	gaggaatgcctctccctcaatgctgtgcccccccgctcgtcgagagggcgacaattctt	420		
Db	307	GAGGACTGCTCTACCTTTACGTTTTCGGCCCTGCTGGCACCAAGCCTGATGCTAAGCTC	366		
QY	421	ccqctcctcgctctacattcacggaggtggctacgcccttcgcg-----gatgcgagcacc	474		
Db	367	CCCGTCATGTTGGATTTACCGTGGTGGCCCTTTGTTGGTTCTTCCTTCCTTACCT	426		
QY	475	ggcagcgactttgcgccttcaccaaagcacacgaggaacaaagatggtgctgtgtaaatctc	534		
Db	427	GGTAACGGCTACGTCACAGGACAGTGTGGAAATGGCCACAGCCTGTGTGTTTGTTCATC	486		
QY	535	cagtcacgtctcggcagcttggtttctcctcgctggccaagccatgaaggaactacgggtga	594		
Db	487	AACTAACGTACCGGCCCTATGGATTCTCGGTGGTGTGTCATCCACCGCTGAGGGTAAC	546		
QY	595	acuaacgcccgttgcctgaccagcaattcgccctcaatgggttcaacagcagctctcg	654		
Db	547	ACCAACGCTGGTCTGCACGACCACGCAAGGCTCTCGAGTGGGTAGCGCAACATTCCT	606		
QY	655	aagttcggcgcaaccccgatcacgttacgatttggggcgagctgcagggcgagggtcc	714		
Db	607	AACTTTGGTGGTGATCCGCACAAGTTCATGATTTTCGGTGAGTCGGCTGGTCCATGAGT	666		
QY	715	gttatgaaccagatcatctgcgaacggcggaacaccgtcaa	755		
Db	667	GTTCCTCACCAAGCTTGTTGCTACGTTGGTGGTGCACACACCTA	707		

RESULT	13	
LOCUS	GCU02541	
DEFINITION	Geotrichum candidum ATCC 34614	DNA linear
ACCESSION	U02541	1635 bp
VERSION	U02541.1	GI:408458
KEYWORDS		
SOURCE		
ORGANISM		
REFERENCE		
AUTHORS	Bertolini, M., Laramee, L., Thomas, D., Cygler, M., Schrag, J. and Vernet, T.	
TITLE	Polymorphism in the lipase genes of Geotrichum candidum strains	
JOURNAL	Unpublished	
REFERENCE	2	(bases 1 to 1635)
AUTHORS	Bertolini, M.	
TITLE	Direct Submission	
JOURNAL	Submitted (14-OCT-1993)	
	Institute - NRCC, 6100 Royalmount Avenue, Montreal, Quebec H4P 2R2, Canada	

FEATURES

source

CDS

Location/Qualifiers

1..1635

/organism="Galactomyces geotrichum"

/strain="ATCC 34614"

/db\_xref="taxon:27317"

<1..1635

/codon\_start=1

/product="lipase II"

/protein\_id="AAA03430.1"

/db\_xref="GI:408469"

/translation="QAPTVLNGNEVISGVVEGKVDTEFKGIPADPLNDLREKHQPQ  
FTGVSYGLKRNDESPACMDQPNSTLIDKALGLAKVPEEREGPLDMAKCTVSMN  
EGLCLVFNAPDEKTRKEDKLPMVNWYIIGAFVYGSSAAIPGNSIYVESINMGQVVFVY  
SNYRTPGPGFGGDRTAEGNTNAGLHQDKLEWSDNIANFGDDPKDKWYIFGESA  
GAMSAVHQIAGGDNTYNGKKLFHSAIIQSGGLPFDHSSVGGPDISYNRFAIYAGC  
DYSASDTECLRSKSSVLDHQAQNSYDLKDLGLLPQFLGFGPDNGNIIIPDAAYE  
RFSRGYAKVYISQNGDEGTAFAFVALNATTHPVKKWLQYIFDASASDRLVLS  
LPQTLSVSGSPFTGLNALTQFKRVAALISDMLEQSPRRVMSATKQVNRWTVYLT  
HLHNLVPFTLGTGHNELIIFQNVNIGPANSYLYRIFSFAHHDPNVGNTLLQWDQVTD  
EKGEMLEIHTMONTDMRTDYEIGISNFTDVNLYG"

BASE COUNT	348 a	460 c	373 g	454 t
ORIGIN				

Query Match 6.2%; Score 107; DB 8; Length 1635;  
Best Local Similarity 56.1%; pred. No. 2.9e-06;  
Matches 225: Conservative 0; Mismatches 170: Indels

[illegible]

RESULT 14

GCU02622	1635 bp	DNA	linear	PLN 22-OCT-1993
LOCUS				
DEFINITION	Geotrichum candidum ATCC 34614	lipase gene, partial cds.		
ACCESSION	U02622			
VERSION	U02622.1	GI:409275		
KEYWORDS				
SOURCE	Galactomyces geotrichum.			
ORGANISM	Galactomyces geotrichum			
	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
	Saccharomycetales; Dipodascaceae; Galactomyces.			
REFERENCE	1 (sites)			
AUTHORS	Bertolini, M., Laramée, L., Thomas, D., Cygler, M., Schrag, J. and Vernet, T.			
TITLE	Polymorphism in the lipase genes of Geotrichum candidum strains			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 1635)			
AUTHORS	Bertolini, M.			
TITLE	Direct Submission			
JOURNAL	Submitted (21-OCT-1993); Bertolini M., Biotechnology Research Institute-NRCC, 6100 Royalmount Avenue, Montreal, H4P 2R2			
	Canada			

FEATURES source

CDS

```

/codon_start=1
/product="lipase"
/protein_id="AAA03435.1"
/db_xref="GI:409276"
/translation="CAPTAVLVNVEISVGLKGVDTFKGIPFADPPVGLRFFKHQPP
FTGSVQGLKANDSSACMLDPNGNAISLLDKVGLGKIIPDLRPPVLDMAQSGSNP
EDCLYLVNFRPAGTPOAKLPVMVIYGGAFVPGSSAYFGNGYVKESVEMGGPVFV
SINRTYTGPGVGLGGDITAEGNTWAGLDORKGLEWYSDNIANFGGDPDKVMLFGE

```

GAMSAVAHQLVAGCDNWTYNGKOLFHSAILQGGGLPDPDSTVSGPESARYQAYAGC  
 DASAGNETLACLRSDYDHSQAYNDLDFGLQFLGFGPRDNGIIPDAAYE  
 LYASAGNAPKYVTIQGDEGFIAPVATNTTPHYVKMLTICSDASIDRLVLS  
 LYPGNSGSEGAFFETGLNALTPQFKRIAAITFDLLFGSPRVLNATKDVNRKTYLAT  
 QLNILVPFFGTTHGSDLLFPQYLVGLGSPSARVYTFISFANHDDNVGNTNLKQMDMTD  
 SGKMIQLMIHMGNSDDETRIEGLTSPESDVTYLFQ#

BASE COUNT  
ORIGIN

Query Match 6.2%; Score 106.8; DB 8; Length 1635;  
Best Local Similarity 56.3%; Pred. No. 3.1e-06;  
Matches 233; Conservative 0; Mismatches 167; Indels 6; Gaps 1;

361	QY	gaggatgctctctcctaattcgttgcgcccgccggtctgacgagggcacaaatttt	420
307	Db	GAGGACTGTCTTACCTTAACGCTTTCCGCCCGCTGGCACCAAGCCTGATGCTAAGCTC	366
421	QY	ccggtcctgctcacattcaaggaggtgctaagccttcggc-----gatcgagacacc	474
367	Db	CCCGTCATGTTTGGAGATTACGGGTGGTGCTTTGTGTTTGGTCTTCTGCTTTTACCCCT	426
475	QY	ggcagcgactttccgccttccaccaagcacacagggaaccaagatggtctgtttaaattctc	534
427	Db	GGTAACGGCTACGTCAAGGAGAGTGTGAAATGGGCCACCGCTGTGTGTTGTTCTCCATC	486
535	QY	cagtaacctgtcggcagcttggtttctctcgtcgtggccaagccatgaaggaactacagtgta	594
487	Db	AACCTACCGTACCGGCCCTATGGATTCTCTGGGTGTGATGCCATCACCGCTGAGGGTAAC	546
595	QY	acgaacgcggcttgttgaccagaatttcgcccttcaatgggttcaacagcacgctctcg	654
547	Db	ACCAACGCTGGTCTGCACGACCAGCGCAAGGGTCTCGAGTGGGTTAGCGACAAACATTGGC	606
655	QY	aagtcgcgggcgaaccccgatacgtttacgatttggcgagctctcagggcgacgggtcc	714
607	Db	AACTTTGGTGGTATCCCGACAAGGTCATGATTTTCGGTGAGTCCCGCTGGTCCCATGACT	666
715	QY	gttatgaaccagatcatttcgaacggcggaacacc	750
667	Db	GTGTCTCACCAAGCTTGTGCTATGTGGTGACAAC	702

RESULT 15

LOCUS	1692 bp	RNA	linear	PAT 29-SEP-1997
E02678	E02678			
DEFINITION	cDNA encoding lipase.			
ACCESSION	E02678			
VERSION	E02678.1	GI:2170906		
KEYWORDS	JP 1990299588-A/1.			
SOURCE	Galactomyces geotrichum.			
ORGANISM	Galactomyces geotrichum.			
	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			

REFERENCE	1 (bases 1 to 1692)	
AUTHORS	Shimada, Y., Tomimaga, Y., Sugihara, A. and Iizumi, T.	
TITLE	GENE OF PROTEIN HAVING LIPASE ACTIVITY	
JOURNAL	Patent: JP 1990299588-A 1 11-DEC-1990;	
	KURITA WATER IND LTD, OSAKA CITY	
COMMENT	OS Geotrichum candidum	
	PN JP 1990299588-A/1	
	PD 11-DEC-1990	
	PF 27-MAR-1989 JP 1989074721	
	PI SHIMADA YUJI, TOMINAGA YOSHIO, SUGIHARA AKIO, IIZUMI TARO PC	
	C12N15/55//C12N9/20,(C12N15/55,C12R1:645);	
	CC strandedness: Double;	
	CC topology: Linear;	
	CC hypothetical: No;	
	CC *source: clone-pgCL1;	
	FH Key	Location/Qualifiers
	FH sig_peptide	1..57
	FT /product= 'signal peptide of lipase', FT	

```

mat_peptide      58..1689
FT               /product='lipase'
FT               1..1692
FT               /product='lipase'.
FT               Location/Qualifiers
                   1..1692
                   /organism="Galactomyces geotrichum"
                   /db_xref="taxon:27317"
BASE COUNT      336 a 466 c 403 g 487 t
ORIGIN

```

Query Match	6.2%	Score 106.8;	DB: 6;	Length 1692;
Best Local Similarity	56.3%;	Pred. No. 3.1e-06;		
Matches 223;	Conservative 0;	Mismatches 167;	Indels 6;	Gaps 1;

  

QY	361	gaggaatgcctctccccaatgctgtgccccgcgcgctgtgcgagggcgacaaattt	420
Db	364	GAGGACTGCTCTACCTTTACGTTTCCGCCCCGCTGGCACCAAGCCTGATGCTAAGCTC	423
QY	421	cccgctcctcgtctacattcacggaggtggcttaoacctctggc-----gatcgagacacc	474
Db	424	CCCGTCATGGTTTGGATTACGGTGGTGCCTTGTGTTTGGTCTTCTGCTCTTACGCT	483
QY	475	ggcagcgacttgcgcgcttccacaagcacacgggaacaaagatggctgttggtaaatctc	534
Db	484	GGTAACGGCTACGTCACAGGAGAGTCTGAAATGGCCAGCCTGTGTGTTTGGTTTCCATC	543
QY	535	cagtaccgctctcggcagcttggtttctcgcctggccaaagcatgaaggactacgggtga	594
Db	544	AACCTACCGTACCGGCCCTATGGATTCTTGGGTGGTATGTCATCACCGCTGAGGGTAAC	603
QY	595	acgaacccggcttgccttgacagcaattcgccttcaatgggttcaacagcagctctcg	654
Db	604	ACCAACGCTGGTCTGCACGACACGCGAAGGGTCTCGAGTGGGTAGCGACAAACATTGCC	663
QY	655	aagttcggcgcaaccgcgcatcacgttaccatttggcgagttcgagggcgaggggcc	714
Db	664	AACTTTGGTGGTATCCCGACAAAGTCATGATTTTCGGTGAGTCCGCTGGTGCATGAGT	723
QY	715	gttatgaaccagatcatttggaaacggcggaacacc	750
Db	724	GTTGCTCACACAGCTGTGTGCTATGGTGGTGACAAAC	759

Search completed: June 29, 2002, 15:03:51  
Job time: 20512 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 29, 2002, 09:17:09 ; Search time 4149.72 seconds  
(without alignments)  
5581.286 Million cell updates/sec

Title: US-09-801-852A-1

Perfect score: 1716

Sequence: 1 atgtcttcaaccttctac.....tctggtagacgggtctc 1716

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*\*

- 1: em\_estba:\*\*
- 2: em\_esthum:\*\*
- 3: em\_estin:\*\*
- 4: em\_estnu:\*\*
- 5: em\_estov:\*\*
- 6: em\_estpl:\*\*
- 7: em\_estro:\*\*
- 8: em\_hc:\*\*
- 9: gb\_estl:\*\*
- 10: gb\_est2:\*\*
- 11: gb\_hc:\*\*
- 12: gb\_gss:\*\*
- 13: em\_gss\_hum:\*\*
- 14: em\_gss\_inv:\*\*
- 15: em\_gss\_pln:\*\*
- 16: em\_gss\_vrt:\*\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	80.2	4.7	620	9	AW790714 D00142-R
2	70	4.1	1800	9	AA415091 MG00026 RC
3	68	4.0	727	10	BI221206 602939553
4	67.8	4.0	625	9	AA202160 LD02542.5
5	67.2	3.9	810	10	BF385648 602047302
6	66.4	3.9	875	10	BF785601 602112390
7	65.8	3.8	783	12	CNS01LTS
8	65	3.8	526	10	BI460262 603201884
9	64.4	3.8	725	10	BG867060 602786057
10	64	3.7	973	10	BI554561 603235901
11	62.6	3.6	1942	11	BC015286 Mus muscu
12	62	3.6	455	9	AI508950 vc28403.Y
13	62	3.6	564	10	BE490305 WHE0359.B
14	62	3.6	676	10	BI564550 RH61852.5
15	62	3.6	730	10	BF505095 AT07260.5
16	62	3.6	1101	12	CNS0027J
17	61.4	3.6	468	9	AW512144 xu54e11.x

18	60.4	3.5	626	9	AI527908
19	59.6	3.5	673	10	BI574042
20	59.2	3.4	649	9	AA246278
21	58.8	3.4	538	10	BF256789 HVSMEF001
22	57.6	3.4	589	10	BG263795 WHE2338.B
23	57.6	3.4	767	10	BG970699
24	57.6	3.4	791	10	BG969645
25	57.2	3.3	568	9	AI516721
26	56.8	3.3	958	12	CNS001Q9
27	56.6	3.3	805	10	BI332322 602983505
28	56.4	3.3	566	10	BI626691 RH67704.5
29	56.4	3.3	959	3	BE636588 rockefell
30	55.6	3.2	700	9	AI527918 uj30904.Y
31	55.4	3.2	383	9	AA230461 MY30811.F
32	55.2	3.2	722	9	AU003170 AU003170
33	55.2	3.2	1770	3	BE636721 rockefell
34	54.8	3.2	242	9	AA377049 EST89551
35	54.8	3.2	477	10	BM374730 EBMa05_SQ
36	54.6	3.2	580	10	BE402096 CSB004D06
37	54.4	3.2	516	9	AI108156 GH06911.5
38	54.4	3.2	526	9	AI108080 GH06811.5
39	54.4	3.2	612	10	BI331296 602982970
40	54.4	3.2	702	10	BI146926 602911553
41	54.4	3.2	716	10	BI220765 602938883
42	54.4	3.2	738	10	BI332320 602983503
43	54.4	3.2	771	10	BI144719 602909509
44	54.4	3.2	778	9	AI062034 GH01076.5
45	54.4	3.2	778	10	BI328564 602985364

#### ALIGNMENTS

RESULT 1

AW790714 620 bp mRNA linear EST 01-MAY-2001  
LOCUS D00142-R Lambda zap, Stratagene Blumeria graminis f. sp. hordei  
DEFINITION cDNA clone D00142 similar to lipase 2 precursor, mRNA sequence.  
ACCESSION AW790714 GI:13902311  
VERSION AW790714.1  
KEYWORDS EST.  
SOURCE Blumeria graminis f. sp. hordei.  
ORGANISM Blumeria graminis f. sp. hordei.  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomyces; Erysiphales; Erysiphaceae; Blumeria.  
REFERENCE 1 (bases 1 to 620)  
AUTHORS Thomas,S.W., Rasmussen,S.W., Glaring,M.A., Rouster,J.A. and Oliver,R.P.  
TITLE Gene identification in the fungal pathogen Blumeria graminis by expressed sequence tag analysis  
JOURNAL Unpublished (2000)  
COMMENT Contact: Rasmussen,S.W.  
Department of Yeast Genetics  
Carlsberg Laboratory  
10 Gl. Carlsbergvej, DK-2500, Copenhagen, Denmark  
Tel: 45 3327 5230  
Fax: 45 3327 4766  
Email: sw@rcr.dk  
POLYA=NO. Location/Qualifiers  
1..620  
/organism="Blumeria graminis f. sp. hordei"  
/db\_xref="taxon:62688"  
/clone="D00142"  
/clone\_lib="Lambda zap, Stratagene"  
/cell\_type="conidia"  
/lab\_host="Hordeum vulgare"

JOURNAL

COMMENT

BASE COUNT

ORIGIN

Query Match

4.7%; Score 80.2; DB 9; Length 620;









```

Db 137 GAGAAATTGCTCTACTTGAATGTGTACCGCGCCCAAGGCAACGGCAGTGCAGTCTCCCTG 196
Qy 421 ccgcctcctctacattcagaggtgctagccttcgagcgtgagcaccgagc 480
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 197 CCCGTAATGTTGTCATACATAGTGGTGGTAATTTCTTCGGTCCGGCATCCACACTG 256
Qy 481 gactttgccccttcacccaagcacacgggaaccaagatggtcgtttgtaaatctccagtac 540
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 257 TACGGCGCGGAGCAATCTTGGCCAC---AAACAAGTGATCCCTCGTCACACTCCAGTAT 313
Qy 541 cgtctcggagcgttggtttcttcctcgtggccaagccatgaaggaactcaggtgtaacaagac 600
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 314 CGTCTCGGTGTGTAGGCTTCTCTTCACGGGGGATGCACACGCTACCGGT-----AAC 367
Qy 601 gccggctgcttgaccagcaattccccccttcaatgggttcaaacagcagctctcgaagtac 660
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 368 TACGGTATGCTTGATCAAGTGTAGCCCTCGATGGTGGTCAATCAGACATCGGCGCCTTC 427
Qy 661 ggccgcaaccccgatcacgttacgatttggggcgagctctgcagcgcaggtccggttatg 720
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 428 GGTGTCATCTCACTCGGTGACACTGTTCGGTGAAGTGGCGCGGTGCTTCGGTGCAG 487
Qy 721 aaccagatcat 731
    ||||| |||||
Db 488 CTGCACATGAT 498

RESULT 8
BI460262 526 bp mRNA linear EST 21-AUG-2001
LOCUS 603201884F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5267757 5',
DEFINITION mRNA sequence.
ACCESSION BI460262
VERSION BI460262.1 GI:15250918
KEYWORDS Est.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 526)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11675 row: g column: 22
High quality sequence stop: 525.
Location/Qualifiers
1..526
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5267757"
/lab_lib="NIH_MGC_97"
/lab_host="DH10B"
/notes="Organ: testis; Vector: pBluescriptR (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcagag
); Oligo-dt primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.2 kb and
normalized to R07 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NHGRI/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
77 a 163 c 187 g 99 t
BASE COUNT
ORIGIN

```

```

Query Match 3.8%; Score 65; DB 10; Length 526;
Best Local Similarity 52.8%; Pred. No. 0.22;
Matches 196; Conservative 0; Mismatches 160; Indels 15; Gaps 2;

Qy 361 gaggattgctctctcctcaatgctgttggcccccgccgctgctgctgagggcgacaaatctt 420
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 123 GAGGACTGTCTGTACCTGAACGTGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 182
Qy 421 ccgcctcctctacattcagaggtggtcagccttcgagcgtgagcaccgagc 480
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 183 CCAGTGTATGTTCTGGTTCCTCCGGGAGGCGCCTTCATCTGGGCGCTGCTTCTTCTGTACGAG 242
Qy 481 gactttgccccttcacccaagcacacgggaaccaagatggtcgtttgtaaatctccagtac 540
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 243 GGCTCTGACTTGGCGCGCGCGCGGAGAAAG-----TGGTGTGTGTGTCTTCGACGAC 293
Qy 541 cgtctcggagcgttgggtttcttcctcgtcgtggccaagccatgaaggaactcaggtgtaacaagac 600
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 294 AGGCTCGGCATCTTCGGCTCTCTGAGCACGACGACACGACGACGACGACGACGACGACGACG 347
Qy 601 gccggctgcttgaccagcaattccccccttcaatgggttcaaacagcagctctcgaagtac 660
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 348 TGGGGGCTGCTGGACAGATGGCGGCTCTCGCTGGGTGGCAGGAGAAATCGGACGCTTC 407
Qy 661 ggccgcaaccccgatcacgttacgatttggggcgagctctgcagcgcaggtccggttatg 720
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 408 GGGGAGACCCAGCAATGTGACCTCTTCGGCCAGTCGGCGGGGCCATGAGCATCTCA 467
Qy 721 aaccagatcat 731
    ||||| |||||
Db 468 GGACTGATGAT 478

RESULT 9
BG867060 725 bp mRNA linear EST 29-MAY-2001
LOCUS 602786057F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4912023 5',
DEFINITION mRNA sequence.
ACCESSION BG867060
VERSION BG867060.1 GI:14217600
KEYWORDS Est.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 725)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10815 row: a column: 16
High quality sequence stop: 724.
Location/Qualifiers
1..725
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4912023"
/lab_lib="NCI_CGAP_SG2"
/lab_host="DH10B (T1 phage-resistant)"
/notes="Organ: salivary gland; Vector: pCMV-Sport6; Site_1:
NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo
dT. Average insert size 1.3 kb. Constructed by Life

```

```

1. 973
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5309856"
/clone_lib="NCI_CGAP_L19"
/lab_host="PH10B (T1 phase-resistant)"
/notes="Organ: liver; Vector: pCMV-SPOK6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dn.
Average insert size 1.9 kb. Constructed by Life
Technologies. Note: this is a NCI CGAP Library."

```

```

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 25 Row: 1 Column: 4
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, Similarity but not identity to protein
This clone has the following problem: frame shifted.
Location/Qualifiers

```









